Mathematical Modeling In Daily Life

桌子一定能放平吗

《红楼梦》前80回和后40回的作者之谜

莎士比亚新诗真伪之鉴定

哥尼斯堡七桥问题与欧拉图

RSA密码体制

股票和股票交易

生活中的数学模型

管宇 主编



浙江工商大學出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

生活中的数学模型


ISBN 978-7-81140-770-9

9 787811 407709 >

定价：29.00元

Mathematical Modeling In Daily Life



主　编　管　宇
副主编　王胜奎　方惠兰　胡海龙



浙江工商大学出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

图书在版编目(CIP)数据

生活中的数学模型 / 管宇主编. —杭州：浙江工商大学出版社，2013.5

ISBN 978-7-81140-770-9

Ⅰ. ①生… Ⅱ. ①管… Ⅲ. ①数学模型 Ⅳ. ①O22

中国版本图书馆 CIP 数据核字(2013)第 088745 号

生活中的数学模型

管　宇 主编

王胜奎　方惠兰　胡海龙 副主编

责任编辑　王黎明
责任校对　何小玲
封面设计　王好驰
责任印制　汪　俊
出版发行　浙江工商大学出版社
(杭州市教工路 198 号　邮政编码 310012)
(E-mail：zjgsupress@163.com)
(网址：http://www.zjgsupress.com)
电话：0571－88904980，88831806(传真)
排　　版　杭州朝曦图文设计有限公司
印　　刷　杭州杭新印务有限公司
开　　本　710mm×1000mm　1/16
印　　张　14.5
字　　数　260 千
版 印 次　2013 年 5 月第 1 版　2013 年 5 月第 1 次印刷
书　　号　ISBN 978-7-81140-770-9
定　　价　29.00 元

前言

模，法式、规范、标准。本义：铸造器物的模子。按《说文》：水曰法、木曰模、土曰型、金曰镕、竹曰范。模型有实物模型和理论模型之分，对描述的对象用数学语言作出的描述和处理都称为数学模型。数学是研究数量、结构、变化以及空间模型等概念的一门科学，它是通过抽象化和逻辑推理，在计数、计算、量度和对物体形状及运动的观察中产生的。数学是现代科学技术的基础，广义地说，任何领域中涉及数量化处理的模型都可以称为数学模型。“万物皆数量”，大至宇宙航天，小到日常生活，数与量无处不在。随着电子计算机科学技术的发展和普及，大规模计算和海量数据的处理已不再是难事。几乎所有领域都会谈及“模型”或“模式”，数学正受到前所未有的重视。但是数学的灵魂是抽象和逻辑，这也使太多太多的人们从小时候喜欢数学，到大学后害怕数学，再到后来逃离数学。

介绍数学的书籍很多，基本上可分为两类，一类是令人生畏的充满定理公式的正宗数学书，一类是科普性读物。后者通常是以智力游戏形式出现，结果是许多人以为数学就是做数字游戏或帮助智力开发。撰写本书的初衷是给人们提供一本介绍数学和数学模型的科普性读物，既不是纯数学，也不是数字游戏。每一章一个主题，从实际背景出发，由浅入深引入数学的思想和方法，试图以不多的篇幅展示量化分析的作用。数学就在我们的身边，数学并不神秘。今天的人们绝对离不开数字产品，如电脑、手机，可是其中又有多少人精通相关软件，又有几人清楚 Windows 或 Word 的运行机制和源代码？即使这样，大家还不都是使用得不亦乐乎！数学完全类似，如果你只关注数学思想、数学方法，而忽视数学公式、数学推导、烦琐计算，会发现数学与你越来越亲近，

你越来越需要它。大多数人就是因为一大堆数学公式、绝对精确推导和一不小心就可能出错的烦琐计算，而敬畏数学、害怕数学、远离数学。如果读者能耐着心浏览本书，并因此而感觉“数学其实很普通，只是因为我太忙没有时间进一步钻研它，否则说不定我也能编出一个某氏定理来”，那我们就非常欣慰了。

本书共 11 章，第 1 章从生活中爬山说起，抛石头测量崖高，先按自由落体运动估算，然后分别增加考虑空气阻力和声音传播，进一步考虑海拔高度、温度、风速等因素影响，尽可能全面完整地回答问题。本章最后给出思考题，并给出相关数据让读者自己完成，这其中介绍了大量的背景知识。如果将正文与思考题内容连起来，就能扩充为 1 个完整的数学模型。第 2 章介绍桌子放平、雨中行走、双层玻璃和连锁经营模型。第 3 章介绍铅球投掷、运动生物力学、体育项目、身体形态特征和赛程安排。第 4 章介绍文学作品和油画辨伪、文物鉴定。第 5 章介绍人口模型和社会历史演变规律。第 6 章介绍饮酒、吃药、打针、传染病的动力学模型。第 7 章介绍战争与博弈。第 8 章介绍图论与网络分析。第 9 章介绍密码学。第 10 章介绍股市分析方法。第 11 章介绍数学建模竞赛和竞赛题目。

本书编写过程中参考了大量的国内外相关文献及网络资料，书后只是列出了其中主要的参考文献。衷心地感谢为数学作出贡献的所有前辈、同仁和数学爱好者们！本书由管宇等人编写，第 6 章由方惠兰执笔，第 8 章由王胜奎执笔，第 9 章由胡海龙执笔，其余各章由管宇执笔，最后由管宇统稿。编者们非常努力地想完成一本令读者满意的科普性读物，因时间和水平限制，难免有这样或那样的错误和欠缺，恳请读者提出宝贵意见，以便进一步修改与完善。特别是有些观点或说法可能早有智者提出，因我们能力所限，书中未提及他们的大名，恳请谅解！

本书的编写得到浙江农林大学教材建设项目经费的资助，本书的出版得到了浙江工商大学出版社的大力支持，在此表示感谢！

管　宇

2012 年 11 月

目录

CONTENTS

第 1 章 从山崖测量说起——什么是数学模型

1.1 山崖测量 …… 1

1.2 数学模型和数学建模 …… 5

1.3 数学模型分类和数学建模过程 …… 7

第 2 章 生活中的数学模型

2.1 桌子一定能放平吗 …… 12

2.2 雨中行走 …… 16

2.3 双层玻璃的功效 …… 20

2.4 连锁经营 …… 26

第 3 章 体育中的数学模型

3.1 铅球投掷 …… 31

3.2 运动生物力学 …… 36

3.3 体育项目与身体形态特征 …… 39

3.4 赛程安排 …… 48

第 4 章 真伪鉴别

4.1 《红楼梦》前 80 回和后 40 回的作者之谜 …… 54

4.2 统计检验 …… 57

4.3 莎士比亚新诗真伪之鉴定 …… 60

4.4 油画造假与鉴定 …… 66

4.5 文物鉴定的科技方法 …… 70
4.6 多管齐下才能更可靠 …… 75

第 5 章 社会历史演变规律——人口模型
5.1 马尔萨斯人口理论 …… 78
5.2 逻辑斯蒂克模型 …… 82
5.3 中国历史人口演变 …… 87
5.4 世界历史人口演变 …… 91
5.5 更精细的人口模型 …… 96

第 6 章 饮酒、吃药和打针——动力学模型
6.1 饮酒驾车 …… 100
6.2 房室模型 …… 103
6.3 饮酒驾车的数学模型 …… 112
6.4 传染病问题 …… 117

第 7 章 战争与和平——博弈理论
7.1 什么是博弈 …… 124
7.2 现代博弈理论 …… 126
7.3 博弈论的基本概念 …… 129
7.4 博弈案例 …… 131

第 8 章 图论与网络分析
8.1 基础知识简介 …… 147
8.2 哥尼斯堡七桥问题与欧拉图 …… 148
8.3 匹配问题 …… 150
8.4 中国邮递员问题 …… 154
8.5 游山玩水问题 …… 158

第9章 密码学及其应用

9.1 密码故事 …… 164
9.2 密码学发展简史 …… 166
9.3 密码学基本知识 …… 168
9.4 代替密码 …… 168
9.5 RSA 密码体制 …… 173
9.6 密码学其他案例 …… 176

第10章 股市实战——预测与决策

10.1 股票和股票交易 …… 182
10.2 波段操作 …… 186
10.3 技术指标分析 …… 189
10.4 旁门左道 …… 194
10.5 预测与决策 …… 198

第11章 数学建模竞赛

11.1 数学建模竞赛 …… 204
11.2 数学建模竞赛题目 …… 207

参考文献 …… 222

第1章 从山崖测量说起——什么是数学模型

1.1 山崖测量

1.1.1 爬山

爬山是一项极佳的有氧运动，能使肌肉获得高出平常10倍的氧气，从而使血液中的蛋白质增多，免疫细胞数量增加，帮助体内的有害物排出；在促进新陈代谢的同时，还可以加快脂肪的消耗，因此爬山也有塑形的功效。爬山是最全面的健身运动，在一步步向上攀爬、跨越时，上下肢反复屈伸，能量消耗极大，使心跳加速，可有效地预防心脑血管病和呼吸系统病，它要比徒步行走和慢跑的运动量大，是一项有效减肥的运动。

山中的空气异常新鲜，对于改善肺通气量、增加肺活量、提高肺的功能很有益处，同时还能增强心脏的收缩能力。山间道路坎坷不平，有益于改善人体的平衡功能，增强四肢的协调能力，尤其是行走在没有经过人为修饰的非台阶路段，可使人体肌纤维增粗、肌肉发达，增强肢体灵活度。保持肌肉和运动器官的协调，增加骨中矿物质的含量，减少骨质疏松的发生，有利于刺激骨细胞的生长，增强韧带和肌腱的力量，防止僵化和早衰。当然，这也可能让膝盖受损。运动学专家建议，上山时身体要前倾，当向上迈的脚踏在台阶上时，后腿应随之用力蹬，而不是简单地起到支撑作用；下山时，为了防止膝关节承受压力增大，前脚向下伸，接触到下一个台阶时，膝盖处应有一定弯曲，前脚掌先着地，再过渡到全脚掌着地，以缓冲膝关节的压力。

另一方面，在山巅之上极目远眺，可以解除眼部肌肉的疲劳，还可使紧张的大脑得到放松和休息，更主要的是此时此刻登山者会获得心情上的愉悦。我们站在山顶，念诵诗神杜甫千古名句《望岳》："岱宗夫如何，齐鲁青未了。造化钟神秀，阴阳割昏晓。荡胸生层云，决眦入归鸟。会当凌绝顶，一览众山小。"此情此景实难用语言表述。身心两相宜，这正是许多人（不管他是达官贵人还是凡夫俗子）只要有时间都愿意去爬山的原因所在吧。

关于登山的著名诗句再摘几首，与读者共享。诗仙李白《独坐敬亭山》：

"众鸟高飞尽,孤云独去闲。相看两不厌,只有敬亭山。"杜牧《山行》:"远上寒山石径斜,白云深处有人家。停车坐爱枫林晚,霜叶红于二月花。"王安石《登飞来峰》:"飞来山上千寻塔,闻说鸡鸣见日升。不畏浮云遮望眼,只缘身在最高层。"毛泽东《为李进同志题所摄庐山仙人洞照》:"暮色苍茫看劲松,乱云飞渡仍从容。天生一个仙人洞,无限风光在险峰。"还有北宋大文豪欧阳修《醉翁亭记》中的千古绝唱:"醉翁之意不在酒,在乎山水之间也。"

爬山时经常会遇到悬崖峭壁,驻足在悬崖绝壁边缘,兴奋、紧张、感叹,给人极大的感官刺激。在享受大自然给人带来的惊奇的同时,你也许出于好奇心,抑或为了展现一下聪明智慧,建议大家测算一下山崖的高度。

1.1.2 山崖测量

例 1.1 假设你站在崖顶,望下万丈深渊,到底有多深呢?大家都会容易想到一种测算方法,那就是在身旁捡一块大些的石头,扔下去听回声返回时间的长短来估算山崖的高度。

显然,首先我们要有计算时间的东西,除了专门的秒表外,平时人们身上带有其他的时间计量工具吗?有,手表、手机。手表可近似到秒,许多非智能手机都有秒表工具。但智能手机却大多没有内置,我们只需上网下载一款手机用秒表小软件即可,这样时间测量问题就解决了。为了后面讨论方便,假定人们用秒表或类似工具能比较准确地测定时间。

接下来,我们自然会想起在中学的物理课中学习过的自由落体运动,在真空环境下任何物体自由下落的重力加速度都是一样的,$g\approx 9.8\ \mathrm{m/s^2}$。假定空气阻力不计,可以直接利用自由落体运动的公式 $h=\frac{1}{2}gt^2$ 来计算高度。我们查阅相关资料,重力加速度主要与纬度和海拔高度有关,数据如下表 1-1。

表 1-1 地球上不同纬度和海拔下的重力加速度($\mathrm{m/s^2}$)

纬度(°)	海拔 (km)			纬度(°)	海拔 (km)		
	0	4	8		0	4	8
0	9.780	9.768	9.756	50	9.811	9.798	9.786
10	9.782	9.770	9.757	60	9.819	9.807	9.794
20	9.786	9.774	9.762	70	9.826	9.814	9.801
30	9.793	9.781	9.768	80	9.831	9.818	9.806
40	9.802	9.789	9.777	90	9.832	9.820	9.807

中国地处北半球,最北端在黑龙江省漠河乌苏里浅滩黑龙江主航道中心

线上(53.56°N),最南端在南海的南沙群岛中的立地暗沙(3.85°N)。北京、上海、广州三大城市市中心纬度分别是39.92°N、34.50°N、23.17°N,海拔高度分别只有31.2、4.5、6.6 m,重力加速度分别是9.80151、9.79460、9.78833 m/s^2。海拔千米以上的省会城市有拉萨(3658 m)、西宁(2261.2 m)、昆明(1891.4 m)、银川(1111.5 m)、贵阳(1071.2 m)、呼和浩特(1063 m)。著名的五岳泰山、华山、衡山、恒山、嵩山主峰海拔分别为1532.7、2154.9、1300.2、2016.1、1491.7 m。除青藏高原外,我国一般山峰海拔多数在一两千米左右,它们的重力加速度约9.78～9.81 m/s^2。因此,我们不妨就用9.8 m/s^2 作为我国多数地区的重力加速度的数值。

假如在山顶放下1块石头让它自由落体,设 $t=4$ s后听到回声,则可求得 $h\approx\frac{1}{2}\times9.8\times4^2=78.4(\mathrm{m})$,即脚下悬崖大概有78.5 m深。如果你是名中学生,能得到这样的结果应该说不错了。也许你会想到应该考虑空气阻力的影响,但不知道怎么算。如果你是名大学生,学过微积分,就能做了。当然,如果你是中学生或者是没学过微积分的人,到此为止不想继续探究。那么我们先来看下面通过相对比较精确的微积分计算得出的结果:62.8 m。78.4与62.8相差15.6,感觉差别有点大,这说明微积分有用,这也从一个侧面告诉大家为什么现在要求绝大多数专业(除艺术、语言之类的少数专业)高校学生一定要学一点高等数学。

现在继续讨论,考虑到除去地球吸引力外,对石块下落影响最大的当属空气阻力。根据流体力学知识,此时可设空气阻力正比于石块下落的速度,阻力系数 K 为常数。牛顿运动学第二定律说,力是产生加速度的原因,物体的加速度跟物体所受的合外力 F 成正比,跟物体的质量成反比,加速度的方向跟合外力的方向相同。因此可得:

$$F=ma=m\frac{\mathrm{d}v}{\mathrm{d}t}=mg-Kv,\text{且 } v(0)=0。$$

这是一个速度 v 关于时间 t 的一阶线性微分方程。令 $k=K/m$,方程简化成

$$\frac{\mathrm{d}v}{\mathrm{d}t}=g-kv,\text{且 } v(0)=0。$$

这是变量可分离微分方程,解得:

$$v=gk^{-1}-gk^{-1}\mathrm{e}^{-kt}。$$

由于速度是距离高度的变化率,即距离高度对时间的导数,对上式等号右边求一次积分得山崖高度的计算公式:

$$h=gk^{-1}(t+k^{-1}\mathrm{e}^{-kt})-gk^{-2}。$$

在互联网上的百度网页中输入"空气阻力系数"一词,会得到相关的风阻系数

资料。垂直平面体风阻系数大约 1.0，球体风阻系数大约 0.5，一般轿车风阻系数 0.28～0.4，好些的跑车在 0.25 左右，赛车可以达到 0.15，飞禽在 0.1～0.2，飞机达到 0.08，目前雨滴的风阻系数最小在 0.05 左右。我们就取 $k=0.05$，$t=4$ s，则可求得 $h\approx 73.4$ m，即脚下悬崖大概有 73.4 m 深。这个结果自然比不考虑空气阻力更接近实际山崖高度。

这里需要说明的是，考虑空气阻力时不必考虑更一般的情形，那么后者的高度计算公式应该包含在前者中，即 $k=0$ 时前者就变成后者。事实上，利用微积分中的泰勒展开式可以验证得：

$$\lim_{k\to 0} gk^{-1}(t+k^{-1}e^{-kt})-gk^{-2}=\frac{1}{2}gt^2。$$

细心的读者可能进一步想到，听到回声再按跑表，计算得到的时间中包含的反应时间应该扣除掉；还有石块落地产生的声音从崖底传到崖顶需要的时间，这个时间也应该扣除。

反应时间是指人从机器或外界获得信息、经过大脑加工分析发出指令到运动器官开始执行动作所需的时间，包括从感觉反应时间（从信息开始刺激到感觉器官有感觉所用时间）到开始动作所用时间（信息加工、决策、发令开始执行所用时间）的总和。由于人的生理、心理因素的限制，人对刺激的反应速度是有限的。一般条件下，人的反应时间约为 0.1～0.5 s，平均 0.2 s；专业运动员通常小于 0.2 s。人的最短反应时间是 0.1 s，因此国际田联规定，反应时间低于 0.1 s 算抢跑，会将运动员罚下场，因为这是压枪。对于复杂现象的选择，人的反应时间达 1～3 s，要进行复杂判断和认识的反应时间平均则达 3～5 s。

由于听到回声再按跑表，是预先设定的事件，反应时间取下限为 0.1 s，假如仍设 $t=4$ s，扣除反应时间后应为 3.9 s，代入求得山崖高度 $h\approx 69.9$ m。

我们都知道声音在空气中传播速度是 340 m/s，当考虑声音传播时间时，令石块下落的真正时间为 t_1，声音传回来的时间记为 t_2，原山崖高度的计算公式修改为：

$$\begin{cases} h=gk^{-1}(t_1+k^{-1}e^{-kt_1})-gk^{-2}, \\ h=340t_2, \\ t_1+t_2=3.9。\end{cases}$$

这是一个非线性方程组，不太容易求得精确解，可用近似计算。

考虑到相对于石块速度，声音速度要快得多，而真实高度与前面计算结果 $h\approx 69.9$ m 的相对误差应该不会太大，完全可用刚才求得的 h 来估算声音传播的时间 $t_2=h/340$，再求石块下落时间 $t_1\approx 3.9-t_2$，然后将 t_1 代入，算出崖高。例如，若 $h=69.9$ m，则 $t_2\approx 69.9/340=0.21$ s，故 $t_1\approx 3.9-t_2=3.69$

s,将 t_1 代入,算出崖高 $h \approx 62.8$ m。

能够想象,62.8 m 虽然不一定就是真实山崖高度,但真实高度与其应该相差无几。当然可以进一步考虑山崖所在的海拔高度,因为不同的海拔高度的加速度是有非常细微区别的;考虑每个人的反应时间,有人反应迅速,有人反应迟钝;考虑山崖处的空气密度,空气密度越大对运动物体的阻力也就越大,风速系数变大,声音传播变慢。但这些影响造成的数值误差显然都不大,而所花的时间却不少,相对于我们仅仅估计山崖的大概高度来说已无此必要。这些进一步的讨论见本章思考题。

山崖测量问题浅显易懂,开始时只需要高中生普通物理运动学知识就能作出初步解答;但要想得到尽可能精确的答案,则要考虑更多影响因素,运用的数学工具也在原来初等数学的基础上增加了一些简单微积分方法。假如你刚才耐心地阅读这一小节,就初步认识了所谓定量地解决实际问题的过程。如果将这一过程进一步抽象,就是所谓的数学建模,其中的主要数学公式就叫数学模型。

1.2 数学模型和数学建模

1.2.1 什么是数学模型和数学建模

数学模型(Mathematical Model)是近年发展起来的新学科,是数学理论与实际问题相结合的一门科学。现在数学模型还没有一个统一的、准确的定义,因为站在不同的角度可以有不同的定义。但基本的内涵是这样的,数学模型是关于部分现实世界和为一种特殊目的而做的一个抽象的、简化的结构。具体来说,数学模型就是为了某种目的,用字母、数学符号建立起来的等式或不等式以及图表、图像、框图、程序等描述客观事物的特征及其内在联系的数学结构表达式,是对实际课题本质属性的抽象而又简洁的刻画。它或能解释某些客观现象,或能预测未来的发展规律,或能为控制某一现象的发展提供某种意义下的最优策略或较好策略。数学建模(Mathematical Modeling)是应用知识从实际课题中抽象、提炼出数学模型的过程。

根据对研究对象观察到的现象及实践经验,它将现实问题归结为相应的数学问题,并在此基础上利用数学的概念、方法和理论进行深入的分析和研究,从而从定性或定量(尤其是定量)的角度来刻画实际问题,并为解决现实问题提供精确的数据或可靠的指导。

数学模型的历史可以追溯到人类开始使用数字的时代,随着人类使用数字,就开始不断地建立各种数学模型,以解决各种各样的实际问题。对大学

生的综合素质测评、对教师工作业绩的评定以及诸如访友、采购等日常活动，都可以建立一个数学模型，确定一个最佳方案。建立数学模型是联系实际问题与数学工具的一座必不可少的桥梁。

1.2.2 建立数学模型的要求

(1)真实完整

真实的、系统的、完整的、形象的客观现象；必须具有代表性；具有外推性，即能得到原型客体的信息，在模型的研究实验中，能得到关于原型客体的原因；必须反映完成基本任务所达到的各种业绩，而且要与实际情况相符合。

(2)简明实用

在建模过程中，要把本质的东西及其关系反映进去，把非本质的、对反映客观真实程度影响不大的东西去掉，使模型在保证一定精确度的条件下，尽可能简单和可操作，数据易于采集。

(3)适应变化

随着有关条件的变化和人们认识的发展，通过相关变量及参数的调整，能很好地适应新情况。

根据研究目的，对所研究的过程和现象(称为现实原型或原型)的主要特征、主要关系采用形式化的数学语言，概括地、近似地表达出来的一种结构。所谓“数学化”，指的就是构造数学模型。通过研究事物的数学模型来认识事物的方法，称为数学模型方法。

1.2.3 数学模型的基本原则

(1)简化原则

现实世界的原型都是具有多因素、多变量、多层次的比较复杂的系统，对原型进行一定的简化即抓住主要矛盾，数学模型应比原型简化，数学模型自身也应是“最简单”的。

(2)可推导原则

由数学模型的研究可以推导出一些确定的结果，如果建立的数学模型在数学上是不可推导的，得不到确定的、可以应用于原型的结果，这个数学模型就是无意义的。

(3)反映性原则

数学模型实际上是人对现实世界的一种反映形式，因此数学模型和现实世界的原型就应有一定的“相似性”，抓住与原型相似的数学表达式或数学理论就是建立数学模型的关键性技巧。

1.3 数学模型分类和数学建模过程

数学模型是数学抽象和概括后的产物，其原型可以是具体对象及其性质、关系，也可以是数学对象及其性质、关系。数学模型有广义和狭义两种解释。广义地说，数学概念如数、集合、向量、方程都可称为数学模型；狭义地说，只有反映特定问题和特定具体事物系统的数学关系结构方可称为数学模型。

1.3.1 数学模型分类

数学模型的种类很多，而且有多种不同的分类方法。

(1)确定性模型和随机模型

描述客体必然现象的模型叫做确定性模型，其数学工具一般是代数方程、微分方程、积分方程和差分方程等。描述客体或然现象的称为随机性模型，随机性模型中变量之间的关系是以统计值或概率分布的形式给出的。在体育实践中常常提到优秀运动员身体体型的数学模型：现代的世界级短跑运动健将模型为身高一米八几、体重 80 公斤左右、100 米成绩 10 秒左右或更好等，但北京和伦敦奥运会冠军牙买加的博尔特则是另类，他身高达 1.96 米。

(2)静态模型和动态模型

静态模型是指要描述的系统各量之间的关系是不随时间的变化而变化的，一般都用代数方程来表达。动态模型是指描述系统各量之间随时间变化而变化的规律的数学表达式，一般用微分方程或差分方程来表示。经典控制理论中常用的系统的传递函数也是动态模型，因为它是从描述系统的微分方程变换而来的。

(3)连续时间模型和离散时间模型

模型中的时间变量是在一定区间内变化的模型称为连续时间模型，各类用微分方程描述的模型都是连续时间模型。在处理集中参数模型时，也可以将时间变量离散化，所获得的模型称为离散时间模型。离散时间模型是用差分方程描述的。

(4)参数模型与非参数模型

用代数方程、微分方程、微分方程组以及传递函数等描述的模型都是参数模型。建立参数模型就在于确定已知模型结构中的各个参数。通过理论分析总是得出参数模型。非参数模型是直接或间接地从实际系统的实验分析中得到的响应，例如通过实验记录到的系统脉冲响应或阶跃响应就是非参

数模型。运用各种系统辨识的方法，可由非参数模型得到参数模型。如果实验前可以决定系统的结构，则通过实验辨识可以直接得到参数模型。

(5)线性模型和非线性模型

线性模型中各量之间的关系是线性的，可以应用叠加原理，即几个不同的输入量同时作用于系统的响应，等于几个输入量单独作用的响应之和。线性模型简单，应用广泛。非线性模型中各量之间的关系不是线性的，不满足叠加原理。在允许的情况下，非线性模型往往可以线性化为线性模型，方法是把非线性模型在工作点邻域内展成泰勒级数，保留一阶项，略去高阶项，即可得到近似的线性模型。

(6)白箱模型、灰箱模型和黑箱模型

根据对某个实际问题了解的深入程度，分为白箱模型、灰箱模型和黑箱模型。白箱模型是指所有过程都建立在因果关系基础上的模型。灰箱模型或概念模型是指那些内部规律尚不十分清楚、在建立和改善模型方面都还不同程度地有许多工作要做的问题。黑箱模型(或称经验模型)是指其内部规律还很少为人们所知的现象。

还可根据建模中所用的数学方法，分为初等模型、微分方程模型、差分方程模型、优化模型等。根据研究课题的实际范畴，分为人口模型、生态系统模型、交通流模型、经济模型、基因模型等。

1.3.2 数学建模过程

建立数学模型需要哪些步骤并没有固定的模式，下面只是按照一般情况，提出一个建立模型的大体过程。

(1)模型的准备

要了解问题的实际背景，明确建立模型的目的，掌握对象的各种信息如统计数据等，弄清实际对象的特征。总之，就是要做好建立模型的准备工作。这一步往往要查阅大量资料，请教专家，以便对问题有透彻的了解。

(2)模型的假设

根据实际对象的特性和建模的目的，对问题进行必要的简化，并且用精确的语言作出假设，是建立模型的第二步，也可以说是关键的一步。有时，假设做得过于详细，试图把复杂的实际现象的各个因素都考虑进去，可能使你很难继续下一步的工作，所以要善于辨别问题的主要和次要方面，抓住主要因素，抛弃次要因素，尽量使问题均匀化、线性化。

(3) 建立模型

根据所做的假设，利用适当的数学工具，建立多个量之间的等式或不等

式关系，列出表格，画出图形，或确定其他数学结构，是建立数学模型的第三步。为了完成这项数学建模的主体工作，人们常常需要具有比较广阔的数学知识，除了微积分、微分方程、线性代数及概率统计等基础知识外，还会用到诸如规划论、排队论、图论、网络及对策论等。推而广之，可以说任何一个数学分支都可能应用于建模过程中。当然，这并不是要求我们对数学的各个分支都精通。事实上，建模时还有一个原则，即尽可能采用简单的数学工具，以便使更多的人能够了解和使用。

(4)模型的求解

对以上建立的模型进行数学上的求解，包括解方程、画图形、证明定理以及逻辑运算等，不仅会用到传统的和近代的数学方法，而且会用到计算机相关的知识。

(5)模型的分析

对上面求得的模型结果进行数学上的分析。有时是根据问题的性质，分析各变量之间的关系和特定性态；有时是根据所得的结果给出数学上的预测；有时则是给出数学上的最优决策或控制。

(6)模型的检验

这一步是把模型分析的结果“翻译”回到实际对象中，如果检验的结果不符合或部分符合实际情况，那么我们必须回到建模之初，修改、补充假设，重新建模，即再按上述步骤操作直到模型检验这一步；如果检验结果与实际情况相符，则进行最后的工作——模型的应用。

整个模型的建立过程可用图 1-1 表示：

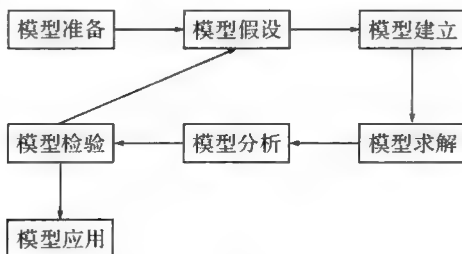


图 1-1　数学建模过程框图

我们看到，数学建模实际上是经过若干次循环而逐渐接近真理的过程。应该指出的是，并非所有的建模过程都要经过上述这些步骤，有时各个步骤之间的界限也并不是那么明显。因此，在建模过程中不要局限于形式上的按部就班，重要的是根据对象的特点和建模的目的，去粗取精、去伪存真，不断完善。

思考题

1. 测量人的反应速度游戏。甲手握一把有刻度的直尺的一端，让直尺垂直下挂，乙站在对面，将拇指和食指对准这把尺子下端（譬如0刻度），但不要接触尺子，并且眼睛要看着甲握尺的手，准备好一旦甲松开手，要尽快用拇指和食指夹住尺子。记录乙夹住尺子处的刻度（注意：在抓住尺子时，手指与尺子接触部分大约1 cm长，应该读取手指下端的刻度值并做好记录），重复5次，求取平均值，这个数值可以反映出你的反应速度的快慢。另外请考虑尺子的长度、重量会否影响测量结果？（提示：反应时间为尺子在自由下落过程中在两指间经过的长度所需的时间）

2. 空气与物体运动相关的问题。理想的自由落体要求在真空环境中进行，现实中是无法达到绝对真空的，空气会影响物体下落速度。空气阻力计算公式如下：

空气阻力＝风阻系数×有效截面积×空气密度×速度平方

其中风阻系数也叫空气阻力系数，是从风洞实验中得来的。具体实验数据如下：垂直平面体风阻系数大约1.0，球体风阻系数大约0.5，一般轿车风阻系数0.28～0.4，好些的跑车在0.25左右，赛车可以达到0.15，飞禽在0.1～0.2，飞机达到0.08，雨滴的风阻系数最小在0.05左右。空气密度定义为，在一个标准大气压下每立方米空气所具有的质量（千克）。它与气温、海拔有关，一般采用的空气密度是指在0℃、绝对标准指标下，密度为1.293 kg/m^3；通常情况下，即20 ℃时取1.205 kg/m^3。换算公式如下：

空气密度＝1.293×（实际压力/标准物理大气压）×（273.15/实际绝对温度）

表1-2　海拔、大气压力和空气密度的关系

海拔高度（m）	大气压力（Pa）	空气密度（kg/m^3）	海拔高度（m）	大气压力（Pa）	空气密度（kg/m^3）
0	101325	1.293	1600	83250	1.063
400	96654	1.233	2000	79489	1.012
800	92072	1.174	2400	75616	0.961
1200	87710	1.114	3000	70097	0.893

利用以上数据，假设我们站在海拔2000m高的山崖上测量，$t=4$s，试估算崖高。

3. 声音传播速度问题。声音借助介质得以传播，真空中就无法传播声音。声音传播公式如下：

$$v=\sqrt{kp/d}=\sqrt{kRT}$$

其中 k 为气体绝热系数，p 为气体压强，d 为气体密度，T 为温度。具体数据(不考虑风速)：空气(15 ℃) 340 m/s，空气(25 ℃)346 m/s，软木 500 m/s，煤油(25 ℃)1324 m/s，蒸馏水(25 ℃)1497 m/s，海水(25 ℃)1531 m/s，铜(棒)3750 m/s，大理石 3810 m/s，铁(棒) 5200 m/s。另外，如果有风速，则空气中实际声速＝静止空气中的声速 ± 风速，顺风时相加，逆风时相减。风速一般每秒几米到十几米，风速 5 m/s 称为微风，可使旗帜展开；8～10.7 m/s 称为清劲风，有树叶、小树枝摇摆；10.8～13.8 m/s 称为强风，大树枝摆动，举伞困难。山谷风的平均速度为每秒 2～4 m，有时可达每秒 7～10 m。假如 $t=4$ s 是在 30 ℃下测得的时间，请估算其中石头落地声音传播到山崖顶的时间。

第 2 章 生活中的数学模型

2.1 桌子一定能放平吗

2.1.1 五花八门的桌子

桌子是一种常用家具，上有平面、下有支柱，通常是表面为光滑平板、由腿或其他支撑物固定起来的家具，可以在上面摆放东西或做事情，如吃饭、写字、工作或玩牌等。康熙字典的解释是“俗呼几案曰桌”，因此“桌”字是现代字，古人叫“案”。桌子的历史起源无法考究，但应该是在金属刀斧工具出现后不久就有了现代桌子的雏形。桌子的外形五花八门，满足不同人的不同需要，主要有方桌、圆桌、开合桌、书桌、电脑桌、茶几、案等。桌子的材料也是多种多样，古代有石材、木材、青铜，近现代增加了玻璃、各种复合材料等。

方桌的桌面为正方形或长方形，且通常是四只脚。一般方桌宽度有 120、90、76 cm 等；长方桌宽度为 80、90、105、120 cm，长度为 150、165、180、210、240 cm 等。其中 76 cm×76 cm 的方桌和 107 cm×76 cm 的长方桌是常用的餐桌尺寸。如果椅子可伸入桌底，即便是很小的角落，也可以放一张六座位的餐桌，用餐时，只要把餐桌拉出一些就可以了。76 cm 的餐桌宽度是标准尺寸，桌宽一般不小于 70 cm，否则，对坐时会因餐桌太窄而互相碰脚。餐桌的脚最好是缩在中间，如果四只脚安排在四角，就有些不方便。中餐桌高度为 75～78 cm，西餐桌高度为 68～72 cm。如桌高 75 cm，则配 41.5 cm 高的座椅。中国传统的八仙桌是宽度至少为 90 cm 的大方桌，每边可坐两人，四边围坐八人（犹如八仙），故民间雅称八仙桌。

圆桌的桌面为圆形或椭圆形，通常前者为三只脚而后者是四只脚。圆桌面直径为 90、120、135、150、180 cm。在一般中小型住宅，如用直径 120 cm 餐桌，常嫌过大，可定做一张直径 114 cm 的圆桌，可坐 8～9 人；平时放 8 张椅子，会显得很拥挤，可放 4～6 张椅子；在人多时，再用折椅，折椅可收藏在贮物室。圆桌象征一家老少团圆、亲密无间，更意味着彼此没有身份高低的差别，能促进一家人沟通与交流，聚拢人气，烘托进食氛围。

开合桌又称伸展式餐桌，可由一张宽度为 90 cm 的方桌或直径为 105 cm 的圆桌变成长度为 135～170 cm 的长桌或椭圆桌(有各种尺寸)，很适合中小型住宅使用。这种餐桌从 15 世纪开始流行，至今已有 500 年的历史，是一种很受欢迎的餐桌。不过要留意它的机械构造，开合时应顺滑平稳，收合时应方便对准闭合。用一张圆餐桌，便可得到较好的空间调整。使用圆桌就餐，还有一个好处，就是坐的人数有较大的宽容度。只要把椅子拉离桌面一点，就可多坐人，不存在使用方桌时坐转角位不方便的弊端。

书桌自然是看书学习时使用的桌子，通常桌面呈长方形。有的学生书桌的桌面做成斜面。单人书桌可用 60 cm×110 cm 的台面，台高 71～75 cm，台面至柜屉底不超过 12.5 cm，否则起身时会撞腿。靠墙书桌，离台面 45 cm 处墙面可设 10 cm 灯槽，上面用作书柜或饰品架，这样，书写时看不见光管，但台面却有充足光照。14 岁以下小孩使用的书桌，台面至少应为 40 cm×50 cm，高度应为 58～71 cm，座椅也要高度合适。

茶几的高度通常为 38～50 cm，以 38 cm 常见。

1971 年以前日本的办公桌标准高度采用 74 cm。由于屡发种种职业病，日本于 1971 年全面修订了办公用具的标准，分别规定了 70 cm 与 67 cm 为男用与女用办公桌的标准高度，从而大大减少了疲劳症状的产生。在英国，目前推荐桌面高度也仅为 71 cm。但在我国，电脑桌、写字台、课桌的高度一般均较高，中国桌子的通常高度都是 76～78 cm，80 cm 的一般都是站立工作用的，比如厨房的地柜，站立使用的家具高度为 80 cm，目前市场上坐着使用的家具(如一般办公桌)的高度都是 76 cm。

事实上，无论是写字台、课桌的高度，还是电脑桌上键盘与鼠标的高度，都应与人在坐姿时的肘部等高或稍低。而显示器的上方也不应高于坐姿的眼睛水平视线，否则将造成视力下降。有关专家建议生产商降低电脑桌的高度，最好能在电脑桌上设计一个折叠式的支肘板。消费者在购买时也可以要求定做尺寸稍低的电脑桌。眼睛与电脑屏幕的适宜距离依个人情况有所不同，但一般以 50～100 cm 为最佳范围。屏幕应旋转在眼睛的俯角 15°为好；光源则应置于上方，以避免反射光直接进入眼球。此外，电脑与环境中的光线对比不宜太大，这样眼睛才不易疲劳。

关于桌子有则趣闻：在 19 世纪末，美国参议院在整修桌子时，工作人员将一些以前的政治家的名字，如韦伯斯特、克莱和卡尔霍恩等，刻在桌子上，这些政治家的确使用过这些桌子。也就是从那个时候起，美国参议员在自己用过的桌子上留名逐渐演变成了一种传统。

2.1.2 桌子能放平吗

现今多为三口之家，为节省空间，方形餐桌平时一般靠墙放置。当有客人时，则将方桌外移至餐厅中央，让就餐空间宽畅。只要有过摆放桌子的经验，大多数人都会有这样一种生活体验：第一次摆放桌子时，很难让四只脚都同时着地，然后经过几次幅度不大的腾挪后，最终摆平桌子。为什么？是不是每次都能摆平桌子？这里面蕴涵着什么样的科学道理？

仔细想想，我们会发现，这里必须先给个约定，这是一张四只脚的桌子。如果是三只脚则无论在什么样的地面上都是能放平的，当然还有其他各种形状各种数目的脚，但四只脚的桌子最常用。如果地面非常平，桌子四只脚的长度完全一样，那么无论怎么摆放都绝对不会出现摇摆的情况。之所以开始时没摆平桌子，显然是要么地面不平要么桌脚长短不一。但为什么通过适当腾挪后却能摆平呢？

设想在一个极端的、凹凸不平的地面上摆放一张方桌，能摆平吗？恐怕多数人会摇头，为什么家里就能摆平？因为家里地面虽不一定能绝对水平但凹凸幅度较小，所以能摆平。下面我们用数学方法严格地论证这一想法。

假设：

(1)方桌(正方形桌)的腿只要有一点接触地面就算着地。桌子四条腿一样长，桌脚与地面接触可视为一个点，四脚的连线呈正方形，即桌子是非常标准的。

(2)地面高度是连续变化的，沿任何方向都不会出现间断(没有像台阶那样的情况)，即地面可视为数学上的连续曲面，即地面虽不是很平坦但也不是很粗糙。

(3)相对于桌脚的间距和桌脚的长度而言，地面是平坦的，因此桌子在任何位置至少有三只脚同时着地，即地面起伏不大。

首先用变量表示桌子的位置，桌脚连线呈正方形，以正方形的中心为对称点，正方形绕中心的旋转正好代表了桌子位置的改变，于是可用旋转角度 θ 这一变量来表示桌子的位置变化。

其次要把桌脚着地用数学符号表示，如果用某个变量表示桌脚与地面的竖直距离，则当这个距离为 0 时，表示桌脚着地了。在桌子挪动改变位置时这个距离是位置变量的函数。

由于正方形的中心对称性，只要设两个距离函数就行了。记 A、C 两脚与地面距离之和为 $f(\theta)$，B、D 两脚与地面距离之和为 $g(\theta)$，显然 $f(\theta)\geqslant 0$，$g(\theta)\geqslant 0$。

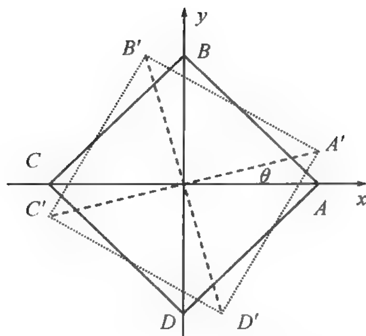


图 2-1 方桌脚面坐标

由假设(2)知：$f(\theta)$，$g(\theta)$都是连续函数。

再由假设(3)知：$f(\theta)$，$g(\theta)$中至少有一个为0。

当$\theta=0$时，不妨设$f(\theta)>0$，$g(\theta)=0$，这样改变桌子的位置使四只脚同时着地，就归结为如下命题。

命题：已知$f(\theta)$，$g(\theta)$是θ的连续函数，对任意θ，$f(\theta)\times g(\theta)=0$，且$f(0)>0$，$g(0)=0$，则存在$\theta_0$，使$f(\theta_0)=g(\theta_0)=0$。

将椅子旋转90°，对角线AC和BD互换。

由$g(0)=0$，$f(0)>0$可知：$f(\frac{\pi}{2})=g(0)=0$，$g(\frac{\pi}{2})=f(0)>0$。

令$h(\theta)=f(\theta)-g(\theta)$，则$h(0)>0$和$h(\frac{\pi}{2})<0$。

由f，g的连续性知：h也是连续函数。

根据连续函数的零点定理，必存在$\theta_0(0<\theta_0<\frac{\pi}{2})$使得$h(\theta_0)=0$，即$f(\theta_0)=g(\theta_0)$。

再由$f(\theta_0)g(\theta_0)=0$，得$f(\theta_0)=g(\theta_0)=0$。

注：模型的巧妙之处在于用一元变量θ表示椅子的位置，以θ的两个函数表示椅子四脚与地面的距离；其中的数学知识只用到连续函数的零点定理或介值定理，相当于大一学生的数学知识水平。利用正方形的中心对称性及旋转90°并不是本质的，读者可进一步考虑四脚呈长方形的情形。

下面的“碰面问题”同样只要利用连续函数的零点定理，就能方便解决。

2.1.3 碰面问题

某日，上午8时张三开车沿沪杭高速从杭州去上海办事，第二天上午8时沿原路从上海返回杭州。则必存在一个地方，张三在两天中的同一时间经过

该地，譬如说两天都是刚好上午 8:55 经过嘉兴服务区。

设 $f(t)$ 和 $g(t)$ 分别表示时刻 t 汽车离杭州的距离，它们是 t 的连续函数，第二天 $T(T>8)$ 时回到杭州。令 $h(t)=g(t)-f(t)$，则 $h(8)>0$，$h(T)<0$，且 $h(t)$ 是连续函数。由零点定理，必存在 $t_0(8<t_0<T)$ 使得 $h(t_0)=0$，即 $f(t_0)=g(t_0)$。

刚才的问题如果换个角度叙述，结论就很显然了。某日上午 8 时，张三与李四两人同时分别从杭州和上海开车，沿沪杭高速分别赶往上海和杭州，则必存在一个地方，两人在同一时间经过该地。更简单明了的说法是张三与李四必在途中相遇。

虽然“连续函数的零点定理”被安排在大学一年级学生的《高等数学》课上学，但实际上，真正理解这个定理根本无需其他知识帮助。打个比方，某人要从五楼下到一楼，不管是坐电梯还是走楼梯，他必然要经过二楼。数学将连续变化的东西抽象成连续函数，日常生活中连续变化的事情太多了，如从出发地出发，汽车、火车都必须先途经中间站点，最后才到达目的地。当然，生活中也有不连续的案例，如某学生成绩很好，直接从二年级跳到四年级，而不读三年级。数学上称不连续的为间断，几乎所有的随时间或其他因素变化的事物都可以分成连续的或间断的两种，读者也完全能够举出生活中许多连续或间断的实例。

用数学语言描述实际问题，可以将问题表达得更本质、更准确。这正是数学模型被越来越多不同领域人士关注的原因所在。所谓模型就是外表貌似不一样的事件，但其中某些数量关系完全一样，因此可用同样的数学方式加以表述，这正是数学的魅力和价值所在。方桌摆放和碰面问题似乎风马牛不相及，但可借用完全一样的“连续函数的零点定理”来解决，这就是数学。

2.2 雨中行走

2.2.1 雨的知识

阳光与水是地球生物赖以生存的两大要素，我们先来认识雨的基本知识。

(1)雨滴的大小

一般情况下，雨滴的直径为 0.5～6 mm，极少数情况下，雨滴的直径会达到 8 mm 甚至 10 mm(据报道，此种情况曾在夏威夷群岛被观测到)。

如果水粒小于 0.5 mm，那么由于大气层上升气流的作用，足以让这等尺寸的水粒留在空中。而雨滴的大小，一般又不会超过 6 mm。由于空气阻力的存在，大个雨滴在下落过程中往往会分解成许多体积骤减的细小雨滴。此

外，这些大个雨滴在下落时不断相互碰撞，也促使它们分崩离析。科学家发现，在实验室条件下，大个雨滴通常在直径达到大约 5 mm 时就开始分化瓦解。在云层中，大个雨滴直径达到 5 mm 时，通常也会因为相互碰撞而纷纷碎裂。这就是为什么在地面上很少看到直径 5 mm 或者直径更大的雨滴，因为雨滴无法避免与许多同类相互碰撞的命运。

(2)雨滴的收尾速度

如果没有空气(阻力)，当天上的云变成雨之后落下来，在经过一路的加速运动之后，到达地面时的速度会达到 300 m/s，这样的速度基本相当于子弹速度的一半。因此若雨滴以此速度下落，别说人了，陆地上根本就不会有活物。由于空气阻力的存在，当雨滴形成并开始下落，经过短暂的变加速(加速度逐渐变小)运动，当雨滴受到的空气阻力等于自身的重力时，雨滴开始进入匀速下落阶段，直至到达地面。物理学上把这个匀速的速度称为物体的收尾速度。

雨滴的收尾速度，与雨滴的大小有较强的相关性。一般而言，毛毛雨(雨滴的直径为 0.5 mm)的收尾速度为 2 m/s，而暴雨(雨滴的最大直径为 5.5 mm左右)的雨滴最大收尾速度为 8～9 m/s。

(3)收尾速度与水柱高度

有了收尾速度，就可以知道雨滴最后的动能，利用初中物理的动能与势能转换公式，就可以解答收尾速度与水柱高度之间的关系。那就是：$\frac{1}{2}mv^2=mgh$。将收尾速度代入此公式，可得：雨滴收尾速度为 8 m/s 时，$h=3.2$ m；雨滴收尾速度为 9 m/s 时，$h=4.05$ m；雨滴收尾速度为 2 m/s 时，$h=0.2$ m。

(4)降雨量

从天空降落到地面上的雨水，未经蒸发、渗透、流失而在水面上积聚的水层深度，称为降雨量(以毫米为单位)，它可以直观地表示降雨的多少。测定降雨量常用的仪器包括量筒和量杯。

(5)雨量等级

小雨：雨点清晰可见，没漂浮现象；下地不四溅；洼地积水很慢；屋上雨声微弱，屋檐只有滴水；12 小时内降水量小于 5 mm 或 24 小时内降水量小于 10 mm 的降雨过程。

中雨：雨落如线，雨滴不易分辨；落硬地四溅；洼地积水较快；屋顶有沙沙雨声；12 小时内降水量 5～15 mm 或 24 小时内降水量 10～25 mm 的降雨过程。

大雨：雨降如倾盆，模糊成片；洼地积水极快；屋顶有哗哗雨声；12 小时内降水量 15～30 mm 或 24 小时内降水量 25～50 mm 的降雨过程。

暴雨:凡24小时内降水量超过50 mm的降雨过程统称为暴雨。根据强度可分为暴雨、大暴雨、特大暴雨三种。暴雨:12小时内降水量30～70 mm或24小时内降水量50～100 mm的降雨过程。大暴雨:12小时内降水量70～140 mm或24小时内降水量100～250 mm的降雨过程。特大暴雨:12小时内降水量大于140 mm或24小时内降水量大于250 mm的降雨过程。

2.2.2 雨中行走

下雨天忘记带伞总是件郁闷的事,因为这样你往往不得不硬着头皮跑回住处,淋一身雨。一个似乎很简单的事情是,你应该在雨中尽可能地快走,以减少雨淋的时间。但如果考虑到降雨方向的变化,在全部距离上尽力地快跑就不一定是最好的策略。所以怎样在跑动中少淋雨,就是生活中非常常见的问题了。

我们的目标:在给定的降雨条件下,设计一个雨中行走的策略,使得你被雨水淋湿的程度最小。

数学工作者在处理实际问题时,总是先找出该问题涉及的主要因素,然后理出这些因素间的数量关系;为表达方便,这些因素就用字母表示,因素间的数量关系则用式子表示;再对式子进行运算,最后得到问题的解。

据此,容易总结出影响雨水淋湿程度的主要因素:淋雨量、降雨的大小、降雨的方向(风)、路程的远近、行走的速度。然后开始以下的数学化处理过程。

(1)符号、说明及假设

把人体视为长方体,身高 h m,宽度 w m,厚度 d m。淋雨总量用 C 升来记。

降雨大小用降雨强度 I mm/h 来描述,降雨强度指单位时间内平面上降下水的厚度。在这里可视其为一常量。

风速保持不变。

用固定奔跑速度 v m/s 跑完全程 D m。

注:任何事物的解决都是在一定的前提条件下进行的。将石头和羽毛在同一高度处下落,谁先着地?石头。如果学过中学物理,都知道自由落体运动,在真空中石头和羽毛下降速度是一样的。也就是说并不是任何时候都是石头比羽毛先着地,因此需要先约定不是在真空中,而是在有空气阻力的地面上。"假设"是让问题的前提条件特殊化,以保证演绎的最后结果的绝对精确;如果没有假设约定,很容易出现一些生活中很少见到的极端反例。

本例假设人体是长方体,目的是简化问题,这样得出的结论是近似的。

如果想要得到精确的结果，这样做显然是不可能的，因为不同人的身材是不相同的，而且真实体形的描述太过复杂最终必然导致无法求解。因此，进行实际问题的数学求解时是离不开"假设"的。

(2)不考虑雨的方向

此时，你的前后左右和上方都将淋雨。

淋雨的面积：$S=2wh+2dh+wd(\mathrm{m}^2)$；

雨中行走的时间：$t=D/v(\mathrm{s})$；

淋雨总量：$C=t\times(I/3600)\times0.001\times S(\mathrm{m}^3)=(D/v)\times(I/3600)\times S(\mathrm{L})$；

其中 D,I,S 为参数，而 v 为变量。

结论：淋雨量与速度成反比。这也验证了尽可能快跑能减少淋雨量的假设。

现在设计具体实例来验证上述公式的合理性。假设一个成年男子：身高 $h=1.7$ m，肩宽 $w=0.4$ m，背厚 $d=0.2$ m。他的淋雨面积 $S=2\times0.4\times1.7+2\times0.2\times1.7+0.4\times0.2=2.12\ \mathrm{m}^2$。

他的百米跑时间为十几秒，不妨取其雨中平均奔跑速度为 5 m/s。假设他需要雨中跑 $D=1000$ m，则雨中时间 $t=1000/5=200$ s。

按大雨的降雨量取 $I=20$ mm/h，身上淋到的总的雨水量：

$C=200\times(20\times0.001/3600)\times2.12=0.002356\ \mathrm{m}^3=2.356\ \mathrm{L}$。

在雨中只跑了 200 秒即 3 分 20 秒，但被淋了大约有 4 酒瓶多的水量，似乎比实际经验明显多了些。这主要是由于不考虑降雨方向使得淋雨面积计算过大造成的。

(3)考虑降雨方向

记雨滴下落速度为 r m/s，方向正对着人前进方向，与人运动方向成夹角 θ。

雨滴的密度为 $p(p\leqslant1)$，表示在一定的时刻在单位体积的空间内，由雨滴所占的空间的比例数，也称为降雨强度系数。则降雨强度 $I=rp$。

因为考虑了降雨的方向，淋湿的部位只有顶部和前面。分两部分计算淋雨量。

顶部的淋雨量：$C_1=$雨中时间×顶部面积×雨滴垂直下落速度$=(D/v)wd(pr\sin\theta)$。

前身的淋雨量：$C_2=$雨中时间×前身面积×雨滴水平运动速度$=(D/v)wh[p(r\cos\theta+v)]$。

总淋雨量：$C=C_1+C_2=(IwD/v)[d\sin\theta+h(\cos\theta+v/r)]$。

取中到大雨的雨滴下落收尾速度为 $r=4$ m/s，其他数值同前，则总淋

雨量：

$$C=(2/900v)[0.2\sin\theta+1.7(\cos\theta+v/4)]。$$

可以看出：淋雨量与降雨的方向和行走的速度有关。问题转化为给定 θ，如何选择 v 使得 C 最小。

情形 1，当 $\theta=90°$ 时，$C=(2/900)[0.2/v+1.7/4]$。结果表明：淋雨量是关于速度的减函数，当速度尽可能大时淋雨量达到最小。假设你以 5 m/s 的速度在雨中跑步，则计算得 $C=0.001033\ m^3=1.033$ L。

情形 2，当 $\theta=60°$ 时，$C=(2/900)[0.425+1.023/v]$。结果表明：淋雨量是关于速度的减函数，当速度尽可能大时淋雨量达到最小。假设你以 5 m/s 的速度在雨中跑步，则计算得 $C=0.0014\ m^3=1.4$ L。

情形 3，当 $90°<\theta<180°$ 时，雨滴将从后面向身上落下，$C=(2/900)[0.425+(0.2\sin\theta+1.7\cos\theta)/v]$。当 $\tan\theta\leqslant-8.5$ 即 $90°<\theta\leqslant96.71°$ 时，淋雨量是关于速度的减函数，当速度尽可能大时淋雨量达到最小。

当 $\tan\theta>-8.5$ 即 $96.71°<\theta\leqslant180°$ 时，淋雨量是关于速度的增函数，不完全符合实际，需另行讨论。首先行走速度越快淋雨时间越短，顶部淋雨量就越少，当行走速度不超过雨滴的水平运动速度时，雨滴将淋在背上；当行走速度恰好等于雨滴的水平运动速度时，淋雨总量最小；当行走速度超过雨滴的水平运动速度时，反而前身会淋到雨，此时淋雨量是关于速度的增函数。所以，当 $96.71°<\theta\leqslant180°$ 时，若行走速度恰好等于雨滴的水平运动速度即 $v=-r\cos\theta$，则淋雨总量最小，仅仅是头顶和肩膀被雨水淋湿。

例如，设 $\theta=120°$，其他数据如前，当 $v=4\sin30°=2$ m/s 时淋雨总量最小，$C=(2/1800)\times0.2\sin120°=0.000192\ m^3=0.192$ L。

(4)结论

若雨是迎着你前进的方向向你落下或接近于 90°的背后落下的，这时的策略很简单，应以最大的速度向前跑；若雨是从你的背后落下的，你应控制你在雨中的行走速度，让它刚好等于落雨速度的水平分量。

2.3 双层玻璃的功效

热水瓶是人们生活的必需品，它能保温，让热水瓶里的热水非常慢地冷却，保温的关键是热水瓶胆。凡弄破过热水瓶胆的人都知道，热水瓶胆是双层玻璃组成的，为什么双层玻璃就能起到保温效果呢？什么样的设计会使保温效果比较好呢？本节先介绍热传导知识，再来认识双层玻璃。

2.3.1 热传导

热量从系统的一部分传到另一部分或由一个系统传到另一个系统的现象叫热传导。热传导是固体中热传递的主要方式，在不流动的液体或气体层中层层传递，在流动情况下往往与对流现象同时发生。热传导实质是大量物质的分子通过热运动互相撞击，使能量从物体的高温部分传至低温部分，或由高温物体传给低温物体的过程。在固体中，热传导的微观过程是：在温度高的部分，晶体中结点上的微粒振动动能较大；在低温部分，微粒振动动能较小；因微粒的振动互相联系，所以在晶体内部就发生微粒的振动，动能由动能大的部分向动能小的部分传递。在固体中热的传导，就是能量的迁移。在液体中热传导表现为：液体分子在温度高的区域热运动比较强，由于液体分子之间存在着相互作用，热运动的能量将逐渐向周围层层传递，引起了热传导现象。气体依靠分子的无规则热运动以及分子间的碰撞，在气体内部发生能量迁移，从而形成宏观上的热量传递。

一般金属都是热的良导体，玻璃、木材、棉毛制品、羽毛、毛皮以及液体和气体都是热的不良导体，石棉的热传导性能极差，常作为绝热材料。

当物体内的温度分布只依赖于一个空间坐标，而且温度分布不随时间的变化而变化时，热量只沿温度降低的一个方向传递，这称为一维定态热传导。此时的热传导可用下式描述：

$$q=-\lambda A(T_2-T_1)/(x_2-x_1)$$

式中 q 为热量通量，T_2-T_1 为温度差，x_2-x_1 为热传递方向的坐标差，λ 为热导率，A 是导热面积。此式表明 q 正比于温度梯度 $\mathrm{d}T/\mathrm{d}x$，但热流方向与温度梯度方向相反。此规律由法国数学家傅里叶于 1822 年首先发现，故称为傅里叶定律。由于热传导的速率与热导率成正比，以窑炉中的换热器为例，通常采用热导率高的材料（如铜、钢、石墨等）作为传热间壁材料。在热绝缘设施中，采用热导率低的材料作为绝热材料。

在最一般的热传导中，温度随时间和三个空间坐标的变化而变化，且伴有热量产生（如反应热）。这时的热传导称为三维非定态热传导，可用下式描述：

$$\frac{\partial T}{\partial t}=\frac{\lambda}{c_p\rho}\left(\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}+\frac{\partial^2 T}{\partial z^2}\right)+\frac{q_c}{c_p\rho}$$

式中 t 为时间，x,y,z 为坐标轴；ρ 为密度，c_p 为定压比热容；$\lambda/(c_p\cdot\rho)$ 为导温系数，又称热扩散系数，表示非定态热传导过程中物体内部温度趋于均匀的能力，即导温系数越大，则温度趋于均匀越快；q_c 为单位体积内热量生成的速率。

在非定态热传导计算中，如果操作是间歇的或周期性的，如蓄热器的操作，这时热传导是非定态的。对于形状简单的物体(如平板、长方体、柱体和圆球体)，可由公式结合一定的初始条件、边界条件求得解析解，但通常求得的解很复杂，往往以无穷级数的形式表示。为便于应用，常将这些结果以图线表述。

对于二维、三维等更复杂的热传导，难以用解析法求解，一般可用数值法求解。

2.3.2 双层玻璃

由于北方冬天气温较低，所以许多建筑物的窗户是双层的，即窗户上装两层厚度为 d 的玻璃夹着一层厚度为 l 的空气，如图 2-1 左图所示。我们要建立一个模型来描述热量通过窗户的热传导(即流失)过程，并将双层玻璃窗与用同样多材料做成的单层玻璃窗(如图 2-1 右图，玻璃厚度为 $2d$)的热量传导进行对比，对双层玻璃窗能够减少多少热量损失给出定量分析结果。

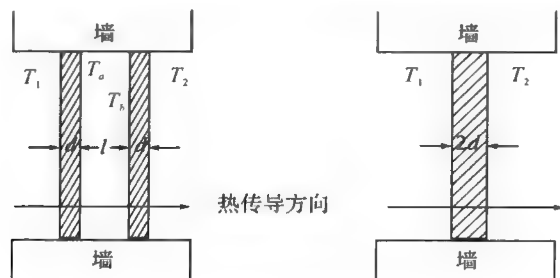


图 2-1 玻璃热传导示意图

(1)模型假设

热量的传播过程只有传导，没有对流，即假定窗户的密封性能很好，两层玻璃之间的空气是不流动的；

室内温度 T_1 和室外温度 T_2 保持不变，热传导过程已处于稳定状态，即沿热传导方向，单位时间通过单位面积的热量是常数；

玻璃材料均匀，热传导系数是常数。

(2)符号说明

T_1—室内温度

T_2—室外温度

d—单层玻璃厚度

l—两层玻璃之间的空气厚度

T_a—内层玻璃的外侧温度

T_b—外层玻璃的内侧温度

k—热传导系数

Q—热量损失

(3)模型建立与求解

由热物理学知道，在上述假设下，热传导过程遵从下面的物理规律：

厚度为 d 的均匀介质，两侧温度差为 ΔT，则单位时间由温度高的一侧向温度低的一侧通过单位面积的热量为 Q，与 ΔT 成正比，与 d 成反比，即

$$Q=k\frac{\Delta T}{d} \tag{1}$$

其中 k 为热传导系数。

①双层玻璃的热量流失

记双层窗内窗玻璃的外侧温度为 T_a，外层玻璃的内侧温度为 T_b，玻璃的热传导系数为 k_1，空气的热传导系数为 k_2，由(1)式可得单位时间单位面积的热量传导(热量流失)为：

$$Q=k_1\frac{T_1-T_a}{d}=k_2\frac{T_a-T_b}{l}=k_1\frac{T_b-T_2}{d} \tag{2}$$

由 $Q=k_1\frac{T_1-T_a}{d}$ 及 $Q=k_1\frac{T_b-T_2}{d}$ 可得 $T_a-T_b=(T_1-T_2)-2\frac{Qd}{k_1}$，

再代入 $Q=k_2\frac{T_a-T_b}{l}$，就将(2)中的 T_a，T_b 消去，变形可得：

$$Q=k_1\frac{(T_1-T_2)}{d(s+2)},s=h\frac{k_1}{k_2},h=\frac{l}{d} \tag{3}$$

②单层玻璃的热量流失

对于厚度为 $2d$ 的单层玻璃窗，容易得出热量流失为：

$$Q'=k_1\frac{T_1-T_2}{2d} \tag{4}$$

③单层玻璃窗和双层玻璃窗热量流失比较

比较(3)(4)有：

$$\frac{Q}{Q'}=\frac{2}{s+2} \tag{5}$$

显然，$Q<Q'$。

为了获得更具体的结果，我们需要 k_1，k_2 的数据，从有关资料可知，不流通、干燥空气的热传导系数 $k_2=2.5\times10^{-4}$[J/(cm·s·℃)]，常用玻璃的热传导系数 $k_1=4\times10^{-3}\sim8\times10^{-3}$[J/(cm·s·℃)]，于是

$$\frac{k_1}{k_2}=16\sim32$$

在分析双层玻璃窗比单层玻璃窗可减少多少热量损失时，我们作最保守的估计，即取 $k_1/k_2=16$，由(3)(5)可得：

$$\frac{Q}{Q'}=\frac{1}{8h+1},h=\frac{l}{d} \tag{6}$$

④模型讨论

Q/Q'的值反映了双层玻璃窗在减少热量损失上的功效，它只与 $h=l/d$ 有关，图 2-2 给出了 $Q/Q'\sim h$ 的曲线，当 h 由 0 增加时，Q/Q'迅速下降，而当 h 超过一定值(比如 $h>4$)后 Q/Q'下降缓慢，可见 h 不宜选得过大。h 大意味着双层玻璃之间的空间大，外观不太美观，隔热效果也增加不明显。

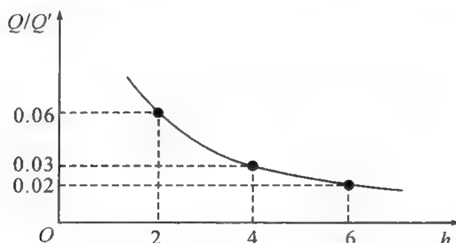


图 2-2　双层玻璃窗与单层玻璃窗热量损失比例曲线

(4)模型的应用

上述模型具有显著的应用价值，制作双层玻璃窗虽然工艺复杂，会增加一些费用，但它减少的热量损失却是相当可观的。通常，建筑规范要求 $h=l/d\approx 4$。按照这个模型，$Q/Q'\approx 3\%$，即双层玻璃窗比用同样多的玻璃材料制成的单层窗节约热量 97%左右。不难发现，之所以有如此高的功效主要是由于层间空气的极低的热传导系数 k_2，而这要求空气是干燥、不流通的。作为模型假设的这个条件在实际环境下当然不可能完全满足，所以实际上双层玻璃窗的功效会比上述结果差一些。

2.3.3　中空玻璃国家标准

由于双层中空玻璃具有隔热隔音效果，国家对于该产品有专门的生产标准。最新的国家标准是《中空玻璃标准》，标准号为 GB/T11944－2002。本标准规定了中空玻璃的规格、技术要求、试验方法、检验规则、包装、标志、运输和贮存要求。本标准适用于建筑、冷藏等用途的中空玻璃。其定义是：中空玻璃(Sealed insulating glass unit)是指“两片或多片玻璃以有效支撑均匀隔开并周边粘接密封，使玻璃层间形成有干燥气体空间的制品”。

详细内容请查阅中华人民共和国国家标准化管理局文件，本书仅列出常用中空玻璃的形状和最大尺寸表。

表 2-1　常用中空玻璃的形状和最大尺寸表

玻璃厚度(mm)	间隔厚度(mm)	长边最大尺寸(mm)	短边最大尺寸(正方形除外,mm)	最大面积(m^2)	正方形边长最大尺寸(mm)
3	6	2110	1270	2.4	1270
	9～12	2110	1270	2.4	1270
4	6	2420	1300	2.86	1300
	9～10	2440	1300	3.17	1300
	12～20	2440	1300	3.17	1300
5	6	3000	1750	4.00	1750
	9～10	3000	1750	4.80	2100
	12～20	3000	1815	5.10	2100
6	6	4270	1980	5.88	2000
	9～10	4550	2280	8.54	2440
	12～20	4550	2440	9.00	2440
10	6	4270	2000	8.54	2440
	9～10	5000	3000	15.00	3000
	12～20	5000	3180	15.90	3250
12	12～20	5000	3180	15.90	3250

2.3.4　热水瓶胆

瓶胆是双层的,内抽真空,使里面的热量不容易被传递出来。胆的里层外围还涂有金属银,热水在瓶胆里,热量在向外流失的过程中又会有一大部分被银镜反射回来从而也能起到保温效果。瓶胆内涂层一般是镀银(银镜),是利用银镜反应镀上去的。现在也有镀铝(铝镜)或离子真空镀膜(钨镜或钛镜)瓶胆。

瓶胆的制作原理如下:工业上平面镜制作及热水瓶胆等玻璃镀银工艺,是利用具还原性的化合物(如醛、单糖、酒石酸钾钠盐等)将银氨配离子还原,使其中的金属银以紧密排列成银箔的方式附着在洁净的玻璃表面形成银镜。由于葡萄糖价格昂贵,目前许多制镜均采用廉价的蔗糖(白砂糖)为原料。一分子蔗糖经酸水解,可以得到两分子均具还原性的单糖(葡萄糖和果糖)。

为了使金属银能在镜面上均匀析出并牢固附着,除了要用酸碱处理、洗涤,使镜面清洁外,还要对镜面进行“敏化”处理。其原因在于,镜面的硅酸钠经酸碱处理时,一部分成为硅酸。当镀银时,银与硅酸交换速度较慢,而镀液

中的碱离子与硅酸交换速度很快，这种活性的差异就影响了镀层的均匀。通常使用 $SnCl_2$ 溶液（敏化液）处理镜面，然后用蒸馏水冲洗洁净后，才可在镜面上开始镀银。

2.4 连锁经营

数学模型讲的是表面不同的实际问题却能用同样的数学方法、数学式子来量化描述，标准化生产是不管在哪里都能用同一个生产经营管理模式生产出同样的产品，这两者有一定的相似性。本节不使用数学公式，只是想通过对连锁经营这样一种非常重要的现代商业模式，说明模式化在生活中是无处不在的。数学是模式化的典范，是模式化的量化表现。

连锁经营是一种商业组织形式和经营制度，是指经营同类商品或服务的若干个企业，以一定的形式组成一个联合体，在整体规划下进行专业化分工，并在分工基础上实施集中化管理，把独立的经营活动组合成整体的规模经营，从而实现规模效益。

2.4.1 连锁经营的特点和优势

连锁经营的最大特点就是统一化，统一店名店貌、统一广告信息、统一进货、统一核算、统一库存和统一管理。统一化带来了许多优势。

(1)连锁经营把分散的经营主体组织起来，具有规模优势。

(2)连锁经营都要建立统一的配送中心，与生产企业或副食品生产基地直接挂钩，节省流通费用，降低成本，一般价格能低于同类商店2%～5%。

(3)连锁经营容易产生定向消费信任或依赖，消费者在任何一个连锁店能得到同样价格、同样服务。

另外，连锁经营特许加盟模式给参与者带来多种好处。

(1)授权人只需投入品牌、经营管理经验等，便可实现规模经营的目的，不仅能在短期内得到回报，而且使无形资产迅速增加。

(2)被授权人由于购买的是已获成功的运营系统，可以省去自主创业不得不经历的一条“学习曲线”，降低了经营风险。

(3)被授权人可以拥有自己的公司，掌握自己的收支，经营启动成本低于其他经营方式，在选址、设计、员工培训、市场等方面，得到经验丰富的授权人的帮助和支持，使其运营迅速走向良性循环。

(4)授权人与被授权人之间不是一种竞争关系，有利于共同扩大市场份额。

2.4.2 连锁经营的结构与形式

(1)连锁经营的结构

由总部、配送中心和若干个门店组成。

统一采购、配送商品,购销分离,统一质量标准,实行规范化经营管理。

总部具备企业法人条件,配送中心可以是总部内设机构,也可是单独设立的配送中心,但由总部控股。

企业用各种规定约束参与者。

(2)连锁经营的形式

连锁经营包括三种形式:直营连锁、特许经营和自由连锁。

①直营连锁:指总公司直接经营连锁店,是由公司本部直接投资和管理各个零售点的经营形态,此连锁形态并无加盟店的存在。

②特许加盟:由拥有技术和管理经验的总部传授经营加盟店的技术经验,并收取一定比例的权利金及指导费,此种契约关系即为特许加盟。特许加盟总部必须拥有一套完整有效的运作技术优势,让加盟店能很快运作起来,同时从中获取利益,加盟网络才能日益壮大。因此,经营技术如何传承是特许经营的关键所在。

③自愿加盟:即自愿加入连锁体系的商店。这种商店由于是原已存在,而非加盟店那样在开店伊始就由连锁总公司辅导创立,所以在名称上应有别于加盟店。

2.4.3 连锁经营的基本前提

连锁经营最早出现在美国,第一家颇具规模的连锁商店是1859年乔治·F·吉尔曼和乔治·亨廷顿·哈特福特在纽约创办的大美国茶叶公司,在6年时间内,该公司发展了26家正规店,全部经销茶叶。20世纪50年代,麦当劳、肯德基引入连锁经营体系,公司得到迅速发展的同时完善了连锁经营业态。连锁经营以其特有的生命力,冲破贸易保护主义的篱笆,从美国向世界各地蔓延。进入21世纪,我国连锁经营爆炸式增长,全面开花,百联集团有限公司、国美电器有限公司、苏宁电器集团、大商集团股份有限公司等迅速成长。

连锁经营的最基本前提是要具备连锁单店复制能力,它很大程度上取决于单店运营的标准化程度,唯有此才不会导致复制走样。国内最成功的餐饮连锁企业,如小肥羊、川味观、真功夫、一茶一坐等,大多是快餐、火锅或休闲餐饮企业,而中式正餐企业比较少,问题就在于中式正餐的标准化难度比较大,而快餐、火锅、休闲餐饮等菜品相对简单,容易实现标准化。有的连锁企

业，在本城市的市场范围内做得好好的，但是一走出去就活不下去，是因为其品牌、产品、服务以及经营管理模式还无法适应其他市场，因此异地扩张能力也是检验连锁企业成长性的重要方面。

2.4.4 兰州牛肉面

"兰州牛肉面"闻名中外，历史悠久，深受普通老百姓的欢迎，全国各地有数不清的"兰州牛肉面馆"。但至今"兰州牛肉面馆"却仍非常简陋，与国内外各种有名的快餐店无法相比。除了兰州本地的牛肉面比较正宗以外，其他地方的牛肉面味道相差很远。卖兰州牛肉面的，一般都是小面馆，地方不大，几张桌子，投资不大，赚钱不多，消费者图的是实惠。兰州牛肉面是连锁经营不成功的案例。

(1)兰州牛肉面历史悠久

有资料记载，兰州牛肉面始于清朝嘉庆年间。相传是甘肃东乡族马六七从河南省怀庆府清化人陈维精处学成而带入兰州的，又经后人陈和声、马保子等人传承创新后，逐渐形成了"一清(汤清)，二白(萝卜白)，三红(辣油红)，四绿(香菜、蒜苗绿)，五黄(面条黄亮)，色、香、味、形俱佳"的特点，是众多牛肉面商家和广大消费者普遍公认的地方特色面食。兰州牛肉面历经百年沧桑而经久不衰，现如今"兰州牛肉面"已成为兰州的一张城市名片和金字招牌，知名度早已远扬海内外。当今兰州牛肉面已成为一种中式大众化快餐面食产业，据兰州市商务局统计，至 2010 年，在兰州市经营牛肉面的餐馆、饭店已达 915 家，其中清真餐馆 639 家，总营业面积 6 万多平方米，从业人员 9500 多人，年营业额达 10 亿元。据粗略统计，国内现有经营兰州牛肉面的餐馆已达 3 万家，从业人员近 34 万人，年营业额达 100 亿元。

(2)兰州牛肉面行业规范化管理的现状

兰州市人民政府、中华人民共和国原国内贸易部、中国烹饪协会和甘肃省商务厅联合，于 2000 年 6 月 18 日至 20 日，在兰州市举办了"首届'金鼎杯'全国兰州牛肉拉面技术比赛"，并结合这一活动，召开了"兰州牛肉拉面与饮食文化发展战略研讨会"，评定兰州牛肉面店(馆)等级，设立牛肉面技术比赛奖。为了加强兰州牛肉面食品质量和等级评定规范化管理，兰州市质量技术监督局、兰州市商业贸易委员会、兰州市第二商业局、兰州市饭店与餐饮协会联合起草，分别于 2000 年 6 月 23 日、2003 年 6 月 11 日经由甘肃省质量技术监督局审定发布了《兰州牛肉拉面》(标准编号 DB62/T685－2000)，《兰州牛肉拉面馆(店)分等定级》(标准编号 DB62/T1031－2003)等两个地方标准。2007 年兰州市还成功地举办了"首届兰州牛肉面节"活动。随着兰州牛肉面

产业规模化、品牌化的逐步形成和发展，涌现出了“金鼎”“马子禄”“黄师傅”等一批被广大消费者普遍认可的牛肉面知名品牌。2003 年 8 月，在省、市政府和有关部门的支持、帮助下，组建成立了兰州市首家牛肉面行业的协会组织——兰州牛肉拉面行业协会。

(3)“兰州牛肉面”集体商标注册中的尴尬

为了依法保护“兰州牛肉面”这一无形资产品牌，兰州牛肉面行业协会委托甘肃恒信商标事务所有限公司，于 2006 年 5 月 25 日向国家商标局提出商标注册申请，分别在第 30 类、第 43 类两个商品和服务类别上，申请注册了 4 件“兰州牛肉拉面”“兰州拉面”集体商标，但是在商标注册实质审查中，国家商标局以不具备集体商标注册条件为由，驳回了兰州牛肉面行业协会提出的这 4 件集体商标注册的申请。世界知识产权组织(WIPO)公布的《商标保护法律协调条约》第 3 条“商标宣言”称：“商标是一种显著的标记，它可以用来或准备用来识别市场上某些商品或服务。”《中华人民共和国商标法》和《集体商标、证明商标注册和管理办法》对集体商标注册和使用管理作了明确的规定，集体商标“是指以团体、协会或者其他组织名义注册，供该组织成员在商事活动中使用，以表明使用者在该组织中的成员资格的标志”。同时，《商标法》又规定，县级以上行政区划名称、公众知晓的外国地名和商品通用名称不得作为商标进行注册，但县以上地名有其他含义或者作为集体商标、证明商标组成部分的除外。将“兰州牛肉拉面”和“兰州拉面”作为集体商标申请注册，其中“兰州”二字作为县以上行政区划名称，与“牛肉拉面”和“拉面”商品通用名称合并组成作为集体商标名称申请注册，就必须按照国家商标局关于含有县以上行政区划名称的地理标志集体商标注册相关规定，提供关于“兰州牛肉拉面”和“兰州拉面”来源于兰州地区的特定质量、信誉和该地区的自然因素、人文因素等特征的历史及现实证明材料；提供兰州市大多数从事兰州牛肉拉面餐饮企业为兰州牛肉拉面行业协会组织成员等相关证明资料。“兰州牛肉拉面”和“兰州拉面”集体商标注册的失败，无疑将对《兰州牛肉拉面》和《兰州牛肉拉面馆(店)分等定级》两个甘肃省地方标准的执行和兰州牛肉面品牌战略的确立及实施带来严重的制约和影响。

(4)无法标准化生产是兰州牛肉面的致命弱点

实际上，成功的快餐连锁企业都离不开标准化生产和特色品牌。

①标准化生产。要求所有连锁企业都出售相同规格、口味的商品，到任何一个连锁店都能吃到正宗的商品。

②特色品牌。虽然大家生产同样的产品，但你的口感要有独特的品味。

麦当劳作为全球最大的快餐连锁店之一，让人们在世界任何一个连锁店

都能吃到一样的食品，当然也可能为了迎合当地民族消费习惯作部分改变。肯德基是与麦当劳的经营产品几乎一样的快餐连锁，但有其不同的一面。肯德基以炸鸡为卖点，而麦当劳以汉堡包为自己的骄傲。

兰州牛肉面至今都没形成统一的品牌，都是些规模小且品牌杂的店面，既不利于对品牌的规范化管理，也不利于形成规模效应。在省内外，不正宗的劣质牛肉面被不断销售出去。这给兰州牛肉面的品牌扩展和宣传带来了比较大的负面影响。

我们去过街边的几个兰州牛肉面馆，感触颇深。几乎每一家的味道都有自己的独特性，但多数是咸得要命，面条经常是粗细不均，这绝不是真正意义上的标准化生产和特色品牌。

思考题

1. 本章论证了正方形桌子通过旋转就能放平，那么长方形桌子(如西餐桌)也一定能放平吗?

2. 存在性证明与构造性证明。只是得出对象一定存在的证明叫做存在性证明，能指出该对象的叫做构造性证明。大数学家希尔伯特曾举例说，“这个班里一定存在一个学生，他的头发数最少，可是我们不可能知道他是谁”，它存在，有时却并不一定容易找到，存在性与构造性问题曾经是20世纪早期科学争议之一，请读者自行查阅相关资料。本章“方桌一定能摆平”用的就是存在性证明，当然如果给出相关的具体数据，还是能把结果给找出来的。假设地面高度：$z=0.02\sin(x^2+y^2)$(m)，方桌边长0.9 m，每脚长0.76 m，现要以坐标点(3, 4)为中心摆放该方桌，请算出方桌能放平时桌子的位置特征。

3. “雨中行走”。假设人是长方体，实际情况明显不是。现在我们考虑一种特殊情形，头是人体中最重要的组成部分之一，下雨时尽量让头部少淋雨是多数人的首选。我们的问题是，如何行走头部淋雨最少?

4. 麦当劳、肯德基在我国连锁经营非常成功，请再至少选取4家在我国进行连锁经营的餐饮企业，查阅它们开店选址、接纳新加盟店的做法，比较分析它们的经营理念和经营模式。

第3章 体育中的数学模型

用现代数学方法研究体育运动是从20世纪70年代开始的。1973年,美国的应用数学家开勒发表了赛跑理论,并用他的理论训练中长跑运动员,使其取得了很好的成绩。几乎与此同时,美国的计算专家艾斯特运用数学、力学理论,并借助计算机,研究了当时铁饼投掷世界冠军的投掷技术,从而提出了他自己的研究理论。而后,艾斯特又据此提出了改进投掷技术的训练措施,从而使这位世界冠军在短期内将成绩提高了4 m,并在一次奥运会的比赛中创造了连破三次世界纪录的辉煌成绩。这些例子说明,数学在体育训练中也在发挥着越来越明显的作用。

3.1 铅球投掷

铅球投掷距离的远近主要取决于三个因素:投掷出手的速度、角度和高度,这三个因素中哪个更重要呢? 为此,我们建立一个数学模型来讨论。在这一模型中不考虑铅球运动员在投掷区域内身体的转动。此外,还给出三条假定:

忽略铅球在运行过程中空气阻力的作用;

投射角度与投射初速度是相互独立的;

将铅球视为一个质点。

3.1.1 初等模型

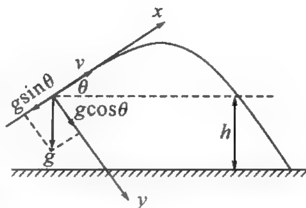


图3-1 沿初速度方向和与初速度方向垂直的方向分解的斜抛运动

设铅球从地平面以初速度 v 和角度 θ 投掷出,出手高度为 h。设铅球的动点坐标为 $(x(t), y(t))$,将铅球的斜抛运动看成是沿初速度方向的匀减速直线运动和与初速度方向垂直的匀加速直线运动的两个分运动的合成(见图3-1),两个方向上的速度分别记为 v_x 和 v_y。利用运动学定律,得

$$\begin{cases} v_{xt}=v-g\sin\theta\cdot t \\ v_{yt}=g\cos\theta\cdot t \\ x=vt-\dfrac{1}{2}g\sin\theta\cdot t^2 \\ y=\dfrac{1}{2}g\cos\theta\cdot t^2 \end{cases}$$

则水平方向的总位移(投掷距离)为 $s_{-}=vt\cos\theta$。

铅球飞行距离的大小取决于 x 方向的分运动，当 x 方向(初速度方向)的分速度为 0 时水平方向的位移也就达到了最大；如果此时的时间恰好就等于铅球飞行的全部时间，那么铅球的射程必将是最大的。

另一方面，将铅球的运动分解成水平方向和竖直方向的运动，建立下列运动学方程

$$\begin{cases} s_{-}=v\cdot\cos\theta\cdot t \\ s_{|}=v\cdot\sin\theta\cdot t-\dfrac{1}{2}gt^2+h \end{cases}$$

由 $v_{xt}=v-g\sin\theta\cdot t=0$ 得，$t=v/(g\sin\theta)$。将此式代入上述运动学方程中的第 2 个方程，且令其中的 $s_{|}=0$，整理解得最佳投掷角度

$$\theta^{*}=\arcsin\sqrt{1/(2+2gh/v^2)}$$

将 θ^{*} 代回，得最远投掷距离

$$s_{-}^{*}=\frac{v}{g}\sqrt{2gh+v^2}$$

如果不考虑出手高度，即 $h=0$，则最佳出手角度 $\theta^{*}=45°$，最大投掷距离为v^2/g。

3.1.2 微积分模型

将铅球的运动分解成水平方向和竖直方向的运动，建立下列运动学方程

$$\begin{cases} s_{-}=v\cdot\cos\theta\cdot t \\ s_{|}=v\cdot\sin\theta\cdot t-\dfrac{1}{2}gt^2+h \end{cases}$$

将 $t=s_{-}/(v\cos\theta)$代入 $s_{|}$ 中得

$$s_{|}=\frac{g}{2v^2\cos^2\theta}s_{-}^2+\tan\theta\cdot s_{-}+h$$

令 $s_{|}=0$，解得落地距离

$$s_{-}=\frac{v^2\sin 2\theta}{2g}\left(1+\sqrt{1+\frac{2gh}{v^2\sin^2\theta}}\right)$$

投掷距离 s_{-} 是关于投掷初速度 v、角度 θ、出手高度 h 的函数，当投掷初速度 v

和出手高度 h 一定时，投掷距离是关于出手角度的一元函数。通过导数求极值的方法便可得到射程 x 为最大值时的 θ 角，即铅球的最佳出手角度。求导并使其等于 0：

$$\frac{\mathrm{d}s_{-}}{\mathrm{d}\theta}=\frac{v^2}{2g}\left[2\cos2\theta\left(1+\sqrt{\frac{2gh}{v^2\sin^2\theta}}\right)-\frac{\cos^2\theta}{\sin^2\theta}\left(1+\frac{2gh}{v^2\sqrt{1+\frac{2gh}{v^2\sin^2\theta}}}\right)\right]=0$$

整理得，$\sin\theta=(2+2gh/v^2)^{-\frac{1}{2}}$。则最佳出手角度为

$$\theta^*=\arcsin\sqrt{1/(2+2gh/v^2)}$$

利用微积分方法与初等方法得到相同结果。

3.1.3 出手高度

刚才将投掷出手的速度、角度和高度，看做完全独立的三个变量。事实上，出手高度 h 会随着出手角度 θ 的变化而变化。通常认为，身体挺直的同时手臂刚好伸直时是最佳出手时机。现在我们把这一因素考虑进去，用 h_1 表示运动员的肩高，b 表示运动员的臂长，则出手高度：$h=h_1+b\sin\theta$。

将铅球的运动分解成水平方向和竖直方向的运动，建立下列运动学方程

$$\begin{cases} s_{-}=v\cdot\cos\theta\cdot t \\ s_{|}=v\cdot\sin\theta\cdot t-\dfrac{1}{2}gt^2+h_1+b\sin\theta \end{cases}$$

将 $t=v/(g\sin\theta)$ 代入 s 中得

$$s_{|}=v^2/g-v^2/(2g\sin^2\theta)+b\sin\theta+h_1$$

令 $s_{|}=0$，得关于 $\sin\theta$ 的一元三次方程

$$2bg\sin^3\theta+2(v^2+h_1g)\sin^2\theta-v^2=0$$

由该方程可解得最佳投掷角度 θ，只是过程非常复杂，略；并据此而得最远投掷距离 s_{-}。

3.1.4 理想数据

借助于数值计算软件如 Matlab，我们能够估算出不同身高的人在不同出手速度下的最佳出手角度、高度和相应的最远投掷距离。

对大多数人而言，站立时其肩高大约是身高的 82%，手臂长约为身高的 52%，兼顾普通人与专业运动员则出手速率取 8.5～15 m/s，身高取 1.5～1.9 m。表 3-1 给出不同身高的人的出手高度在随出手角度的变化而变化的条件下，出手速度、最佳出手角度、出手高度和最大射程的理想数据。

表 3-1　不同身高的人的出手高度在随出手角度的变化而变化的条件下，出手速度、最佳出手角度、出手高度和最大射程

出手速度 v(m/s)	身高 H(m)	1.5	1.6	1.7	1.8	1.9
	肩高 h_1(m)	1.23	1.312	1.394	1.476	1.558
	手臂长 b(m)	0.78	0.832	0.884	0.936	0.988
8.5	出手高度 h(m)	1.73	1.84	1.94	2.06	2.17
	出手角度 θ(°)	39.53	39.28	38.99	38.69	38.42
	投掷距离 s(m)	8.93	9.01	9.11	9.21	9.3
10	出手高度 h(m)	1.74	1.85	1.97	2.08	2.19
	出手角度 θ(°)	40.81	40.58	40.35	40.13	39.91
	投掷距离 s(m)	11.81	11.91	12.01	12.1	12.2
12	出手高度 h(m)	1.76	1.87	1.99	2.11	2.22
	出手角度 θ(°)	41.94	41.76	41.59	41.41	41.24
	投掷距离 s(m)	16.35	16.46	16.56	16.66	16.76
14	出手高度 h(m)	1.76	1.87	1.99	2.11	2.22
	出手角度 θ(°)	42.68	42.54	42.4	42.27	42.13
	投掷距离 s(m)	21.69	21.8	21.9	22	22.11

由表 3-1 知，当出手速度一样时，不同身高的人最远投掷距离相差不大，相关不超过 1 m；真正影响距离的是出手速度，这也正是专业运动员与普通人之间的区别。

本模型没有考虑空气阻力、风速之类的因素。考虑到国际比赛的标准铅球为球形，表面光滑；男子铅球直径为 11～13 cm，重量 7.26 kg；女子铅球直径为 9.5～11 cm，重量 4 kg；空气阻力、风速的影响应该相对较小，几乎可以忽略不计。一些论著和教科书中提到的铅球出手角度的推荐值为 38°～42°。

3.1.5　实际情形

1989 年李梅素以 21.32 m 的成绩打破亚洲纪录时，铅球的出手角度为 36.4°。如果李梅素在推铅球时，铅球的出手角度提高 2°～3°，她推铅球的成绩是不是会有更为令人惊喜的提高呢？

(1)出手角度对铅球飞行距离的影响

一般来讲，运动员推铅球时最能发挥身体肌肉力量的动作是躯干的横向扭转，躯干角度越小越容易发挥身体力量，这与 38°～42°的最佳铅球出手角度

是矛盾的。而片面地追求最佳的铅球出手角度或者最好的身体肌肉用力角度并不是推铅球的目的,推铅球的目的是增加投掷距离,是技术动作的结果。在铅球出手角度和身体肌肉用力角度中,不管单独地追求哪一项都显然是不可能达到理想目的的。要取得最好的投掷成绩,只有对铅球出手角度与身体肌肉用力角度进行最佳组合,也就是追求这两项的合成效果。

李梅素推铅球的特点是水平用力距离较长,躯干扭转发力动作速度快,所以对她来说,提高铅球出手角度也许没有带来推铅球成绩的增长,而是下降。统计表明,铅球出手角度有下降趋势。有的学者认为,铅球出手角度下降的主要原因是,运动员要通过适当降低铅球出手角度来换取能获得更大利益的身体力量的充分发挥。铅球的最佳出手角度(37°~41°)小于45°,是由于出手点(一般为2.2~2.3 m高)高于落地点。出手高度增加,最佳出手角度下降,反之亦然。

(2)出手高度对铅球飞行距离的影响

铅球的出手高度取决于运动员的身高和手臂长度以及运动员的身体和训练技术水平。出手高度增加多少,会引起铅球飞行距离差不多相同的增加。从运动实践的角度看,每个运动员的出手高度和出手角度是相对恒定的,不会因外界因素而出现明显的改变。

(3)出手速度对铅球飞行距离的影响

在推铅球的滑步阶段和最后用力阶段,速度矢量指向不同方向,如果这种指向缺乏直线性,则会影响铅球的出手速度。铅球的出手速度是推铅球运动各阶段铅球获得速度的总和。速度矢量的方向遵循平行四边形法则。推铅球中最重要的因素是出手速度。出手初速度增加150%,铅球飞行距离是原来的2.25倍。高水平的铅球运动员原地推19~20 m相应的出手速度为13 m/s,滑步结束时铅球的速度为2.0~2.5 m/s,如果运动员能够将两个速度进行数学结合(速度的方向一致)完成推铅球动作,那么,铅球的出手速度可能是15~15.5 m/s,相应的距离是25~26 m。而较轻的铅球(男6 kg,女3 kg)出手速度为16 m/s。然而由于滑步阶段和最后用力阶段铅球的速度缺乏直线性,这就导致了其末速度与初速度相比有相当大的损失(速度损失达60%~70%)。

(4)身体环节的协调配合对铅球飞行距离的影响

铅球的运动及其速度是身体环节——腿、躯干和手臂共同活动及其速度合成的结果。我们假设有一些方法可以使手和铅球获得最大速度。第一种是,最大速度可能在身体所有环节达到速度最大时获得。根据生物力学的“特殊冲量相等”原理,躯干和手臂速度最大值的瞬间一致,可以使出手速度

增加。如果在躯干和下肢处于最大速度的情况下，产生手臂的最大伸展速度将导致铅球出手速度的下降，反之亦然；躯干与下肢产生最大速度间的时间越长，铅球的总速度越小。但是这些环节的力及其运动模式的最佳结合很难确定，还有待进一步研究证实。

3.2 运动生物力学

3.2.1 什么是运动生物力学

运动生物力学是应用力学原理和方法研究生物体的外在机械运动的生物力学分支。狭义的运动生物力学研究体育运动中人体的运动规律。按照力学观点，人体或一般生物体的运动是神经系统、肌肉系统和骨骼系统协同工作的结果。神经系统控制肌肉系统，产生对骨骼系统的作用力以完成各种机械动作。运动生物力学的任务是研究人体或一般生物体在外界力和内部受控的肌力作用下的机械运动规律，它不讨论神经、肌肉和骨骼系统的内部机制，这些内部机制属于神经生理学、软组织力学和骨力学的研究范畴(生物固体力学)。在运动生物力学中，神经系统的控制和反馈过程以简明的控制规律代替，肌肉活动简化为受控的力矩发生器，作为研究对象的人体模型可以忽略肌肉变形对质量分布的影响，简化为由多个刚性环节组成的多刚体系统。相邻环节之间以关节相连接，在受控的肌力作用下产生围绕关节的相对转动，并影响系统的整体运动。

对于人体运动的研究最早可追溯到15世纪达·芬奇在力学和解剖学基础上对人体运动器官的形态和机能的解释。18世纪已出现对猫在空中转体现象的实验和理论研究。运动生物力学作为一门学科是20世纪60年代在体育运动、计算技术和实验技术蓬勃发展的推动下形成的。20世纪70年代，H.哈兹将人体的神经—肌肉—骨骼大系统作为研究对象，利用复杂的数学模型进行数值计算，以解释最基本的实验现象。T. R.凯恩将描述人体运动的坐标区分为内变量和外变量，前者描述肢体的相对运动，为可控变量；后者描述人体的整体运动，由动力学方程确定。这种简化的研究方法有可能将力学原理直接用于人体实际运动的仿真和理论分析。由于生物体存在个体之间的差异，实验研究在运动生物力学中占有特殊且重要的地位。实验运动生物力学利用高速摄影和计算机解析、光电计时器、加速度计、关节角变化、肌电仪和测力台等工具测量人体运动过程中各环节的运动学参数以及外力和内力的变化规律。

在实践中，运动生物力学主要用于确定各专项体育运动的技术原理，作

为运动员的技术诊断和改进训练方法的理论依据。此外,运动生物力学在运动创伤的防治、运动和康复器械的改进、仿生机械(如步行机器人)的设计等方面也有重要作用。同时还为运动员的选材提供了依据。

3.2.2 运动生物力学在武术运动中的应用

本小节介绍牛顿三大定律在武术运动中的应用。

(1)牛顿第一定律在武术运动中的应用

牛顿第一定律即惯性定律,其内容为:任何物体在不受外力作用或所受合外力为零时,将保持其原有的静止状态或匀速直线运动状态。我们在生活中经常有这样的体会,要使静止的载重推车启动比较困难,要花很大力气,然而,一旦推动起来,车子就依靠惯性继续运动,只要花很小的力气就可以保持运动的状态。同理,根据牛顿第一定律,运动时应特别注意动作的连贯性,尽可能避免频繁停止动作然后再重新启动,以减少不必要的能量损耗。惯性大小是由物体的质量所决定的。“克服惯性”实际上是改变物体的运动速度(包括大小和方向)。

从体育运动的动作系统来讲,保持动作的连贯和系统运动的速度,要依靠惯性。例如,对于长拳套路演练中连续的跳跃动作腾空飞脚接旋飞脚组合而言,要使动作连贯、协调、省力,就必须考虑到惯性在其中所起的作用。腾空飞脚经过助跑、踏跳、腾空完成动作后,落地时,人体有向前、向下运动的趋势,对于下一动作旋飞脚来说,即具有向前、向下的初速度。由惯性定律知,保持一定的速度比改变速度要容易、省力得多。此时,运动员应利用人体运动的惯性,继续上前两步加速,接做旋飞脚。这样可以减少能量的损耗,并且还利用了腾空飞脚完成落地时的末速度,使旋飞脚踏跳前的初速度更大,从而能保证更高质量地完成旋飞脚动作。反之,如果腾空飞脚完成落地后,运动员通过肌肉用力制动,将身体速度停止为零,然后再上步加速接做旋飞脚,那么,两个动作中间衔接过程的能量消耗是不必要的浪费,并且旋飞脚从零速度启动加速,在踏跳前的初速度也较小,人体具有的动量就小,无形中增加了完成动作的难度,影响了动作完成的质量。

在对抗性项目,如太极推手中,运用惯性定律的时机也很多。惯性定律认为物体有保持其运动状态的特性。太极拳技击理论要求“闪开正中,定横中”,较形象地说明了技击对抗中对惯性的应用。当对方直线进攻我胸部时,对方具有一定的向前移动的速度,如果此时出乎对方意料,我用闪空法躲过对方的进攻,对方就会由于惯性失去平衡而跌倒,或者在闪空的同时从对方后背沿对方运动方向加一力,如此效果会比前一种更明显,这就是所谓的“顺

水推舟”。

惯性定律在武术运动中的应用比较广泛，总的来说有这样几种情况。

①在套路演练中利用自身的惯性使用“巧劲”。惯性定律在套路演练中的应用主要表现在动作之间的衔接上，后一动作应利用前一动作结束时的速度，运用惯性定律继续做后一动作，前一动作结束时的末速度即是后一动作开始时的初速度。套路演练时，合理利用惯性定律，可以减少不必要的能量消耗，节约体能，同时还可以使动作连贯流畅，省力协调，收放适度，演练风格快慢相间，节奏明快。

②在对抗性项目中利用对方的惯性掌握进攻时机和方式。对抗性项目主要利用惯性定律的人体有保持本身运动状态的特性。在对方进攻时，把握对方招数已发但还不能收回之机施以各种招法，或闪让，使对方由于自身的惯性作用失去平衡而跌倒；或趁对方处于运动状态重心还未稳定的时机，在与对方运动方向一致或稍有夹角的方向上加一力，使其失去重心，从而起到事半功倍的效果。

在武术运动中合理地利用惯性，既可以省力、省能量，又可以使动作连贯、圆活、美观，增强动作效果，是借力打力的基本原理之一。

(2)牛顿第二定律在武术运动中的应用

要使物体运动状态发生改变，必须对物体施加力。牛顿第二定律揭示了物体运动与作用其上的力的关系：当一个物体受到的合外力不为零时，物体运动的加速度与合外力成正比，与其质量成反比。加速度的方向与合外力的方向一致。用公式可表示为：

$$F_1+F_2+\cdots+F_l=ma$$

其中 $F_1+F_2+\cdots+F_l$ 表示作用在物体上的合外力矢量，m 表示物体的质量，a 表示物体质心运动的加速度矢量。

散打和拳击比赛中，为使参赛运动员的竞技能力相当，差别不悬殊，规定将运动员按体重(质量)划分为不同的级别。因为人体惯性的大小与质量有关。质量大的运动员不灵活，但惯性大，稳定性好，抗击打能力强；质量小的运动员灵活性好，但稳定性差，抗击打能力差。并且，根据牛顿第二定律，在合外力一定时，质量越大的运动员产生的加速度越小，改变其运动状态越慢，即闪躲慢、出招慢、变化招法也慢；而质量小的运动员产生的加速度大，运动速度和运动状态发生改变得快，在对抗时表现为出招快、变化灵活。

(3)牛顿第三定律在武术运动中的应用

牛顿第三定律表明了力是物体间的相互作用。若甲物体以力 F_1 作用于乙物体，则乙物体会同时以力 F_2 作用于甲物体，F_1 与 F_2 大小相等、方向相

反，且沿同一直线作用在两个不同的物体乙、甲上，用公式表示：

$$F_1 = -F_2$$

F_1、F_2 为矢量，负号表示两力方向相反。例如，运动员打沙袋，运动员拳头以力 F_1 打到沙袋上，同时运动员的拳头受到沙袋所给的反作用力 F_2 与 F_1 大小相等、方向相反，并且 F_1 与 F_2 作用在同一直线上。但是 F_1 作用在沙袋上而 F_2 作用在拳头上。俗话说“要想打人，先学会挨打”，原理就是牛顿第三定律——作用力与反作用力定律。

只有外力才能引起整个人体力学系统运动状态的变化，但是人体可以通过自身的内力主动作用于外界物体，则根据牛顿第三定律，外界物体会同时给人体一个反作用力，这一反作用力即是外力。外力引起整个人体力学系统运动状态的变化。

另有一典故，说杨氏太极宗师杨露禅，将鸟放于掌中，鸟在掌中振翅却不能飞。其原因可以从牛顿第三定律作用力与反作用力的角度来分析。鸟要起飞，需要有外界给予的向上的力 F，产生向上的加速度，这样才能使向上的速度 v 不断增大，至飞起。这个向上的外力 F，通过鸟腿主动用力蹬杨大师的掌心而由杨大师的掌心产生对鸟向上的反作用力实现。当鸟向下蹬掌心时，掌心并不是不动的，掌心通过向下移动，使鸟脚的蹬地力不能作用到手掌上，也就不能获得反作用力。对于人体也一样，只有外力才能引起人体整个力学系统总质心的变化，并且，人体主动获得外力的途径是：只有通过肌肉主动收缩产生内力，内力作用于外界人体或物体，利用牛顿第三定律，外界人体或物体同时给人体施加一反作用力，对人体来说即获得了外力。

3.3 体育项目与身体形态特征

3.3.1 身体形态特征

(1)身体形态特征定义

身体形态是指人体外部与内部的形态特征，在一定程度上反映着相应的生长发育水平、身体机能水平和竞技能力水平。身体形态是运动员选材的重要指标，也是培养优秀运动员的前提。随着现代竞技运动的迅速发展和运动训练的科学化进程，科学选材的问题也越来越受到人们的广泛关注。每个项目都有符合该项目特点的系统而科学的形态指标，这些指标也成为选拔运动员的一个外在条件。

通过前人的研究发现，某一专项运动的优秀运动员的体格、体型具有一定的规律性，运动员通过合理的运动训练可以使自己的身体形态达到该项目

技术要求的最理想状态，从而充分发挥自己的潜能，取得好的运动成绩。譬如游泳运动员通常手臂明显偏长，美国天才游泳冠军菲尔普斯是典型案例。

(2)身体形态特征包含测量数据

①体格测量包括：重量测量、长度测量、宽度测量和围度测量。指标包括：体重、身高、上肢长、上臂长、下肢长 B、下肢长 A、下肢长 C、大腿长、小腿长+足高、肩宽、骨盆宽、胸围、上臂紧张围与上臂放松围、腰围、大腿围、小腿围。

②身体成分测量包括：体脂(百分比)、去脂体重、脂肪、矿物质、肌肉、蛋白质、体液和基础代谢率。

③体型测量包括：身高、体重、上臂围、小腿围、肱骨内外上髁间距、股骨内外侧髁间距、髂嵴上部皮脂厚度、上臂部皮脂厚度、肩胛下角皮脂厚度、小腿内侧皮脂厚度。

④生长发育评价：克托莱指数(体重/身高)×1000、身高胸围指数(胸围/身高)×100、利维指数(体重$^{1/3}$/身高×1000)。

本书不是体育方面的专著，身体形态特征指标主要涉及最常见的身高、体重和克托莱指数。本章仅介绍篮球和中长跑运动员的部分成果，其他项目可查阅相关文献。

(3)克托莱指数

克托莱指数亦称“体重—身高指数”或“肥胖指数”，是人体测量的复合指标之一。

19 世纪中叶，比利时人克托莱在长期研究后提出著名的克托莱指数：体重(kg)/身高(cm)×1000，单位为克/厘米(g/cm)。通过体重与身高的比例关系表示每厘米身高的重量，以相对体重或等长体重反映人体围、宽、厚度及机体组织密度，在古人类学和人体发育评价研究中广泛运用。同性别、同年龄的指数随身高的增长而增大；人体生长发育中，男女性指数均随年龄的增长而增大，男到 21 岁、女到 19 岁基本稳定。男女指数均值曲线在 11 岁和 15 岁时交叉，11～15 岁间女性指数大于男性，11 岁前、15 岁后女性指数小于男性。

3.3.2 男子篮球运动员身体形态特征

随着现代男子篮球运动技术、战术的迅速发展和日趋完善以及队伍大型化、运动员位置模糊化和球队攻防整体协同化的趋势，对篮球运动员的身体条件和各项身体素质提出了更高要求。身体素质是技术的基础，身体形态是运动员发展的先决条件。本小节介绍顶尖男子篮球运动员的身体形态特征。

(1)研究对象

研究对象为第29届奥运会男子篮球比赛前8名队96名运动员的身高和体重的基本参数,按1至8名顺序排列,分别是美国队、西班牙队、阿根廷队、立陶宛队、希腊队、克罗地亚队、澳大利亚队和中国队,每队都有12名球员,共96人。

对奥运会男子篮球比赛前8名队不同位置运动员的身体形态与8强整体水平、对中国队与其他球队不同位置运动员的身体形态特征进行比较分析。

(2)运动员身高、体重和克托莱指数特征的比较分析

运动员身高、体重和克托莱指数的基本统计数据见表3-2。

表3-2　第29届奥运会男子篮球比赛前8名队运动员身体形态基本统计量的相关参数一览表

	指标	前8	美国	西班牙	阿根廷	立陶宛	希腊	克罗地亚	澳大利亚	中国
身高(cm)	平均	201.3	199.4	200.0	199.8	201.6	203.6	202.9	201.8	201.7
	标准差	8.7	8.2	10.2	7.9	6.7	7.2	8.8	10.0	11.5
	最小值	180.0	183.0	182.0	185.0	193.0	193.0	185.0	180.0	185.0
	最大值	226.0	211.0	215.0	210.0	212.0	216.0	217.0	213.0	226.0
	排序		8	6	7	5	1	2	3	4
体重(kg)	平均	100.4	100.0	98.7	103.2	99.3	108.8	97.2	98.4	97.5
	标准差	14.3	11.7	16.6	15.9	8.1	21.6	11.3	12.3	12.8
	最小值	73.0	77.0	77.0	81.0	87.0	93.0	83.0	75.0	73.0
	最大值	170.0	120.0	127.0	140.0	110.0	170.0	111.0	118.0	125.0
	排序		3	5	2	4	1	8	6	7
克托莱指数	平均	497.2	500.4	491.0	514.7	491.9	533.2	478.1	486.4	481.9
	标准差	55.9	43.3	60.5	61.6	28.2	95.1	42.1	41.9	40.2
	最小值	394.6	420.8	403.1	424.1	432.8	476.0	412.9	416.7	394.6
	最大值	817.3	568.7	590.7	666.7	532.0	817.3	533.7	554.0	553.1
	排序		3	5	2	4	1	8	6	7

①运动员平均身高。当今篮球运动发展趋势是“更高、更强、更快”,预示着高大运动员越来越多,比赛强度越来越大,对抗越来越激烈。第29届奥运会男子篮球比赛前8名队运动员整体平均身高为201.3 cm,希腊、克罗地亚、澳大利亚、中国和立陶宛队运动员的平均身高高于整体平均身高,西班牙、阿根廷和美国队运动员的平均身高低于整体平均身高。中国队运动员在身高

上具有一定的优势。

②运动员平均体重。第29届奥运会男子篮球比赛前8名队运动员整体平均体重为100.4 kg，希腊和阿根廷队运动员的平均体重高于整体平均体重，美国、立陶宛、西班牙、澳大利亚、中国和克罗地亚队运动员的平均体重低于整体平均体重。中国队运动员的平均体重为97.5 kg，在8支队伍中排序第7，说明中国队运动员的身体充实度和力量与其他队相比处于劣势。根据运动员平均身高特征分析，前8名球队中除希腊队外，其他球队运动员并未表现出身高越高、体重越大的发展趋势，说明运动员个体差异较为明显。

③运动员平均克托莱指数。第29届奥运会男子篮球比赛前8名队运动员整体平均克托莱指数为497.2，希腊、阿根廷和美国队运动员的平均克托莱指数高于整体水平，立陶宛、西班牙、澳大利亚、中国和克罗地亚队运动员的平均克托莱指数低于整体水平。中国队运动员的平均克托莱指数为481.9，在8支队伍中排序第7，说明中国队运动员的身体充实度和身体质量与其他球队比较处于劣势，也就是说，中国队运动员在激烈对抗的比赛中身体对抗能力处于劣势。

(3)不同位置运动员身高、体重和克托莱指数特征的比较分析(表3-3)

表3-3　第29届奥运会男子篮球比赛前8名队不同位置运动员身体形态基本统计量的相关参数一览表

指标		美国	西班牙	阿根廷	立陶宛	希腊	克罗地亚	澳大利亚	中国
中锋	身高(cm)	209.5	215.0	207.0	209.8	211.0	211.5	213.0	220.0
	排序	7	2	8	1	5	4	3	1
	体重(kg)	112.0	122.5	126.0	107.8	130.8	110.8	115	112.5
	排序	6	3	2	8	1	7	4	5
	克托莱指数	534.4	569.8	607.8	513.7	620.4	523.8	539.9	510.2
	排序	5	3	2	7	1	6	4	8
前锋	身高(cm)	204.5	205	204.6	200.0	205.3	205.3	205.3	204.5
	排序	6	4	5	8	1	2	3	7
	体重(kg)	105.8	107.3	108.2	97.0	101.0	101.0	101.0	100.3
	排序	3	2	1	8	4	5	6	7
	克托莱指数	517.2	523.0	528.7	484.9	492.0	492.0	492.0	490.5
	排序	3	2	1	8	4	5	6	7

续表

		美国	西班牙	阿根廷	立陶宛	希腊	克罗地亚	澳大利亚	中国
后卫	身高(cm)	192.7	191.7	192	195	196.6	195.4	192	193.6
	排序	5	8	6	3	1	2	7	4
	体重(kg)	92.2	85	89	93	96	89	87.3	87.6
	排序	3	8	5	2	1	4	7	6
	克托莱指数	477.9	443.4	463.4	476.9	488.3	455.9	453.5	451.7
	排序	2	8	4	3	1	5	6	7

①中锋运动员身高、体重和克托莱指数特征的比较分析。

第29届奥运会男子篮球比赛前8名队中锋运动员整体平均身高为211.7 cm,中国、西班牙和澳大利亚队中锋运动员的平均身高高于整体水平,克罗地亚、希腊、立陶宛、美国和阿根廷队中锋运动员的平均身高低于整体水平。中国队比中锋运动员整体平均身高高8.3 cm,说明在比赛中中国队中锋运动员与其他球队中锋运动员相比有很好的空中优势。

第29届奥运会男子篮球比赛前8名队中锋运动员整体平均体重为117kg,希腊、阿根廷和西班牙队中锋运动员的平均体重高于整体水平,澳大利亚、中国、美国、克罗地亚和立陶宛队中锋运动员的平均体重低于整体水平。中国队在8强中排序第5,比整体平均体重低4.5kg。说明中国队中锋运动员横向发育水平与纵向高度相比要差,在激烈对抗的比赛中,对抗能力影响高度优势的发挥。

第29届奥运会男子篮球比赛前8名队中锋运动员整体平均克托莱指数为552.5,希腊、阿根廷和西班牙队中锋运动员的平均克托莱指数高于整体水平,澳大利亚、美国、克罗地亚、立陶宛和中国队中锋运动员的平均克托莱指数低于整体水平。中国队在8强中最低,比整体平均克托莱指数低42.3。说明中国队中锋运动员在激烈对抗的比赛中,对抗能力处于劣势。

②前锋运动员身高、体重和克托莱指数特征的比较分析。

前8名队前锋运动员整体平均身高为203.8 cm,希腊、克罗地亚、澳大利亚、西班牙、阿根廷、美国和中国队前锋运动员的平均身高高于整体水平,立陶宛队前锋运动员的平均身高低于整体水平。中国队在8强中排序第7,比整体平均身高略高0.7 cm。说明中国队前锋运动员与其他球队的前锋运动员相比,空中优势不明显。

前8名队前锋运动员整体平均体重为102.1 kg,阿根廷、西班牙和美国队前锋运动员的平均体重高于整体水平,希腊、克罗地亚、澳大利亚、中国和立

陶宛队前锋运动员的平均体重低于整体水平。除大于前锋整体水平的3个球队外，其他球队差距都不是很大。中国队排序第7，比整体平均体重低1.8 kg。中国队的前锋运动员在激烈对抗的比赛中，对抗能力略处劣势。

前8名队前锋运动员整体平均克托莱指数为500.9，阿根廷、西班牙和美国队前锋运动员的平均克托莱指数高于整体水平，希腊、克罗地亚、澳大利亚、中国和立陶宛队前锋运动员的平均克托莱指数低于整体水平。中国队排序第7，比整体平均克托莱指数低10.5。说明中国队前锋运动员在激烈对抗的比赛中，对抗能力处于劣势。

③后卫运动员身高、体重和克托莱指数特征的比较分析。

前8名队后卫运动员整体平均身高为193.6 cm，希腊、克罗地亚、立陶宛和中国队后卫运动员的平均身高高于整体水平，美国、阿根廷、澳大利亚和西班牙队后卫运动员的平均身高低于整体水平。中国队排序第4，刚好是整体水平。说明中国队后卫运动员在激烈对抗的比赛中，略有空中优势。

前8名队后卫运动员整体平均体重为89.8 kg，希腊、立陶宛和美国队后卫运动员的平均体重高于整体水平，克罗地亚、阿根廷、中国、澳大利亚和西班牙队后卫运动员的平均体重低于整体水平。中国队排序第6，比整体平均体重低2.2 kg。说明中国队后卫运动员在激烈对抗的比赛中，对抗能力与其他球队后卫运动员比较略处劣势。

前8名队后卫运动员整体平均克托莱指数为463.6，希腊、美国和立陶宛队后卫运动员的平均克托莱指数高于整体水平，阿根廷、克罗地亚、澳大利亚、中国和西班牙队后卫运动员的平均克托莱指数低于整体水平。中国队排序第7，比整体平均克托莱指数低11.9。中国队后卫运动员在激烈对抗的比赛中，对抗能力与8强其他球队后卫运动员比较略处劣势。

(4)结论

从以上的比较分析看，运动员并未表现出身高越高体重就越大的身体形态特征，运动员个体差异较为明显，但体重与克托莱指数保持着相同的变化趋势，即运动员体重越大，克托莱指数就越高。

3.3.3 优秀男女中长跑运动员身体形态特征

现在，我们对1983年至2002年世界大赛成绩前120名的优秀运动员（共169名世界优秀中长跑运动员）的身体形态与机能特征进行研究，运算数据见表3-4至表3-9。

表 3-4　中外优秀女子中长跑运动员克托莱指数平均值比较

项目	中国			外国		
	身高(cm)	体重(kg)	克托莱指数	身高(cm)	体重(kg)	克托莱指数
800 m	166.8±4.3	52.6±4.5	315±20.9	170.5±4.2	54.9±3.9	332±19.1
1500 m	166.8±4.1	53.4±3.9	314±22.2	166.6±6.4	53.5±3.5	321±16.1
5000 m	164.3±2.2	51.9±2.8	315±15.7	165.4±5.5	51.8±3.1	313±19.3
10000 m	164.3±2.2	51.9±2.8	315±15.7	165.4±5.5	51.8±3.1	313±19.3
马拉松	160.1±5.6	48.7±5.7	285±25.3	162.0±4.2	47.6±5.5	294±16.4

表 3-5　中外优秀男子中长跑运动员克托莱指数平均值比较

项目	中国			外国		
	身高(cm)	体重(kg)	克托莱指数	身高(cm)	体重(kg)	克托莱指数
800 m	177.1±3.0	65.6±2.3	371±13.9	183.8±4.8	69.0±4.7	371±19.3
1500 m	177.0±3.3	63.1±5.5	356±26.7	177.3±7.8	63.0±6.5	354±23.9
5000 m	173.8±3.6	57.1±3.5	329±17.8	177.3±4.4	62.9±3.7	355±14.0
10000 m	171.3±4.7	55.8±3.5	326±17.2	173.3±4.3	59.9±4.1	345±29.4
马拉松	171.4±6.7	55.3±3.5	333±39.9	175.9±6.5	60.4±6.1	343±30.3
3000 m 障碍	174.9±3.8	60.5±3.7	346±18.8	179.1±2.5	66.3±3.8	370±20.0

表 3-6　中国优秀中长跑运动员躯干与四肢指数平均值比较

性别	项目	指距－身高指数	胸围/身高指数	100×髂宽/肩宽	100×髂宽/髋宽	100×踝围/跟腱长
男子	800 m	4.71±3.4	50.63±8.5	69.76±3.7	86.73±3.9	100.5±1.46
	1500 m	3.59±4.1	51.45±6.6	69.76±3.7	86.73±3.9	100.5±1.46
	5000 m	3.53±3.6	50.96±7.4	71.40±3.9	88.90±4.2	99.80±1.86
	10000 m	3.53±3.6	50.96±7.4	71.40±3.90	88.90±4.2	—
	马拉松	1.49±2.2	51.96±1.6	71.40±3.90	88.90±4.2	—
	3000 m 障碍	3.58±3.5	51.70±0.4	71.72+3.8	88.60±3.6	102.5±1.49
女子	800 m	2.72±3.2	49.29±1.5	75.30±3.5	88.70±3.7	100.8±1.44
	1500 m	1.50±4.0	49.87±1.7	75.30±3.5	88.70±3.7	100.8±1.44
	5000 m	－0.33±3.4	—	76.90±2.5	89.90±5.9	100.7±1.51
	10000 m	－0.33±3.4	—	76.90±2.5	88.90±5.9	100.7±1.23

表 3-7　中国优秀中长跑运动员身体成分指数比较

性别	项目	脂肪(%)	去脂体重(kg)	去脂体重/体重指数	去脂体重/身高指数
男子	800 m	10.18±1.12	56.56±4.21	89.82±1.12	32.27±2.05
	1500 m	10.18±1.12	56.56±4.21	89.82±1.12	32.27±2.05
	5000 m	10.30±1.13	54.88±3.91	89.70±1.03	31.60±1.70
	10000 m	10.30±1.13	54.88±3.91	89.70±1.03	31.60±1.70
	3000 m 障碍	10.30±1.13	54.88±3.91	89.70±1.03	31.60±1.70
	马拉松	10.30±1.13	54.88±3.91	89.70±1.03	31.60±1.70
	竞走	10.30±1.13	54.88±3.91	89.70±1.03	31.60±1.70
女子	800 m	15.20±2.49	43.67±3.36	84.80±2.49	26.52±1.56
	1500 m	15.20±2.49	43.67±3.36	84.80±2.49	26.52±1.56
	5000 m	15.20±2.49	43.67±3.36	84.80±2.49	26.52±1.56
	10000 m	15.20±2.49	43.67±3.36	84.80±2.49	26.52±1.56
	马拉松	15.20±2.49	43.67±3.36	84.80±2.49	26.52±1.56
	竞走	15.20±2.49	43.67±3.36	84.80±2.49	26.52±1.56

表 3-8　中长跑运动员机能代谢特征指数比较

项目	运动时间(min)	ATP-CP 供能(%)		无氧糖酵解供能(%)		有氧供能(%)	
		男子	女子	男子	女子	男子	女子
800 m	1.5～2	30	30	65	65	5	5
1500 m	3～4	20	20	55	55	25	25
5000 m	10～16	5	5	15	15	80	80
10000 m	30～50	5	5	5	5	90	90
马拉松	135～180	0	0	5	5	95	95
竞走	135～180	0	0	5	5	95	95

表 3-9 中国优秀中长跑运动员心肺功能指数比较

项目	肺活量(mL)		肺活量/体重(mL/kg)		最大摄氧量相对值(L/min)	
	男子	女子	男子	女子	男子	女子
800 m	4400±496	3668±298	66.5±5.5	69.7±4.4	69.20±7.79	76.10±6.1
1500 m	4320±631	3480±384	66.1±7.5	68.1±5.6	69.20±7.79	76.10±6.1
5000 m	4433±484	3553±396	67.5±5.4	67.8±5.9	72.55±3.47	—
10000 m	4433±484	3553±396	67.5±5.4	67.8±5.9	72.55±3.47	—
马拉松	4514±348	3553±396	75.7±3.3	—	68.18±6.29	—
3000 m 障碍	4725±497	—	72.2±4.9	—	72.55±3.47	—

分析结果如下。

①世界优秀男女中长跑运动员的身高、体重和克托莱指数变化反映出，随着竞赛距离的增加三项指数呈现出递减的变化趋势，而且男女运动员的身体形态在整体水平上也表现出胸围相对较小、四肢相对较短的瘦、窄、小专项形态特征。

②随着比赛距离的增加，男女运动员的 100×髂宽/肩宽指数以及 100×髂宽/髋宽指数均呈现出递增趋势，这一变化特征可能与长距离和超长距离运动时间较长，而且对躯干、腰腹和髂、髋周围肌群力量耐力要求较高有关。虽然女子运动员的 100×髂宽/肩宽指数比男子运动员要大一些，但是它和 100×髂宽/髋宽指数一样都没有表现出明显的专项和性别差异，而是呈现出相对稳定的髂、髋比例特征。

③与短跑或直道跨栏运动员相比，中长跑运动员的 100×踝围/跟腱长指数明显要小，跟腱相对较短，这可能与中长跑需要长时间持续运动而不需要较强的爆发力有关，也可能与运动员的先天遗传因素以及需要承受较大运动量和消耗较多机体能量有关。而中长跑运动员所呈现出的身体轻、体内肌肉含量较少和肌肉不发达的体成分特征，则可能与运动员先天遗传因素、需要承受较大负荷量以及消耗较多的能量有关。

④中跑运动所需能源大约有 65%是由无氧糖酵解提供的，长跑运动所需能源有 80%以上是靠有氧代谢提供的。因此，中跑运动训练应在适度发展磷酸原供能水平基础上突出发展无氧糖酵解供能水平。而长跑运动训练则应在重点发展有氧代谢能力基础上，适度发展无氧糖酵解供能水平。

⑤女子中跑运动员的肺活量指数和肺活量/体重指数均高于长跑和超长跑运动员，但男子却没有体现出这一特征，即使是最大摄氧量相对值，男子运动员也呈现出无规则分布状态。而且男女运动员的肺活量指数、肺活量/体

重指数、最大摄氧量相对值与竞技水平之间并不存在线性关系，而是体现出明显的个体差异。

3.4 赛程安排

无论是体育比赛还是各种文化娱乐比赛，人们总希望从赛程安排、现场评分到成绩排名都是完全公平公正的。本节通过最基本的单循环赛程安排问题，向读者展示其中的数学思想。

3.4.1 单循环赛程安排问题

要想最后成绩合理，循环赛是首选。所谓循环赛就是参加对象两两都要进行比赛。先来考虑最简单的单循环赛程安排问题：设有 n 支球队在同一场地上进行单循环赛，共要进行 $n(n-1)/2$ 场比赛，问如何安排赛程使对各队来说尽量公平。

3.4.2 假设和说明

(1)模型的假设

①假设每个队的实力相当；

②在比赛过程中不会出现意外的停赛事故；

③给每支球队都编上队号，依次为 $1,2,3,\cdots,n$；

④每两场比赛间隔时间相等，且这段时间不足以恢复队员体力。

(2)符号说明

n：球队号($n=1,2,3,\cdots,n$)；

N：n 个队参赛时，共需比赛的场次数；

$SUP(n)$：相隔场次数上限；

$d(v_i)$：第 i 支球队的度数，即第 i 支球队已参赛场数；

r：平均相隔场次数；

r_{max}：r 的上限；

f：总体最大偏差；

f_{min}：f 的下限；

g：球队最大偏差；

g_{min}：g 的下限；

R：所有球队组成的一个集合；

E：所有球队比赛的集合；

E_1：已安排比赛的场次的集合；

$e(v_i,v_j)$：第 i 支球队与第 j 支球队比赛。

3.4.3　5 支球队赛程安排

现有 5 支球队，给出各队每两场比赛间至少相隔一场的赛程，5 支球队共比赛 $N=10$ 场。以 5 支球队为顶点，每两支球队之间进行的比赛为相应顶点间的边，则构成一个 5 阶完全图 G，在图 G 中寻找一条满足条件的路径即可，见图 3-2 和表 3-10。

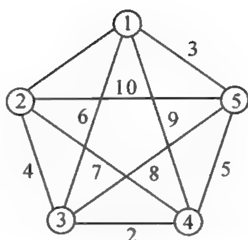


图 3-2　5 支球队比赛关系图

表 3-10　5 支球队比赛赛程安排表

	1	2	3	4	5	每两场比赛间相隔场次数
1		1	6	9	3	1　2　2
2	1		4	7	10	2　2　2
3	6	4		2	8	1　1　1
4	9	7	2		5	2　1　1
5	3	10	8	5		1　2　1

3.4.4　n 支球队赛程安排

当 n 支球队比赛时，各队每两场比赛中间相隔的场次数的上限是多少？分 3 种情况考虑。

(1)第 1 种情况

单独考虑某一个队的最大上限，球队可以连续打比赛，最大上限为 $SUP(n)=n(n-1)/2$，结果很明显。

(2)第 2 种情况

若只单独考虑某一个队的最大上限，每支球队的连续两场比赛间至少相隔一场。

公理 1　如果 $2\leqslant n\leqslant 4$，则必有一支球队连续的两场比赛之间不相隔任何场次。

定理 1　如果 $n=5$，则各队每两场比赛中间相隔的场次数的上限是 2。

证明略。

定理 2　如果 $n\geqslant 6$，则各队每两场比赛中间相隔的场次数的上限为 $SUP(n)=n(n-1)/2-2n+4=n(n-5)/2+4$。

证明:用数学归纳法来证明。

当 $n=6$ 时,$SUP(6)=7=6(6-5)/2+4$。

当 $n=7$ 时,$SUP(7)=11=7(7-5)/2+4$。

假设 $n=k(k>7)$ 时,$SUP(k)=k(k-1)/2-2k+4=k(k-5)/2+4$ 成立。

由于是以 k 个球队为顶点,以每两支球队之间的比赛为边,构成一个 k 阶完全图。设 v_i 能达到上限 $SUP(k)$,即 v_i 打第一场和第 $SUP(k)+2$ 场。此时 $d(v_i)=2$,随后的比赛就间隔一场打一场,直至最后一场比赛。当 $n=k+1$ 时,就新加一个顶点 v_{k+1}。要使 v_i 打完第 $SUP(k+1)+2$ 场时 $d(v_i)=2$,则 $d(v_{k+1})$ 应为 $k-2$,所以 $SUP(k+1)=SUP(k)+k-2$。因此,当 $n=k+1$ 时,$SUP(k+1)=k(k-1)/2-2k+4+k-2=k(k+1)/2-2(k+1)+4$。即当 $n=k+1$ 时成立。定理得证。

(3)第 3 种情况

从整体考虑各队的最小上限,当 n 支球队比赛时,各队每两场比赛中间相隔的场次数的上限 $r\leqslant(n-3)/2$。

证明:设 n 为奇数,$n=2k+1$。共比赛 $N=k(2k+1)$ 场。考察前 $k+1$ 场,有 $2k+2$ 个队参赛,于是至少有 1 个队两次参赛,这个队在这两场比赛间相隔场次数 r 不超过 $(k+1)-1-1=k-1=(n-3)/2$。

设 n 为偶数,$n=2k$,共比赛 $N=k(2k-1)$ 场。同上,在前 $k+1$ 场中至少有 1 个队(记这样的一个队为 A)两次参赛,记 A 第 j 场比赛在赛程中是第 a_j 场,于是 $a_1\geqslant1,a_2\leqslant k+1$。

若 $a_2\leqslant k+1$,则 $r=a_2-a_1-1\leqslant k-2=(n-3)/2$;

若 $a_2=k+1$,但 $a_1>1$,则同样地 $r=a_2-a_1-1\leqslant k-2=(n-3)/2$;

若 $a_1=1,a_2=k+1$,在前 $k+1$ 场中除 A 外有 $2k$ 个队参赛,于是至少又有 1 个队(记这样的一个队为 B)两次参赛。记 B 第 j 场比赛在赛程中是第 b_j 场,则必有 $b_1\geqslant1,b_2<k+1$,或 $b_1>1,b_2\leqslant k+1$(即不可能 $b_1=1,b_2=k+1$ 同时满足),故 $r=b_2-b_1-1\leqslant k-2=(n-3)/2$。

①偶数支球队赛程安排。以 $n=8$ 为例,赛程安排如表 3-11。同理,可安排任意偶数支球队的赛程。

表 3-11　8 支队参赛的比赛顺序

第一轮	第二轮	第三轮	第四轮	第五轮	第六轮	第七轮
(1)1—8	(5)1—7	(9)1—6	(13)1—5	(17)1—4	(21)1—3	(25)1—2
(2)2 -7	(6)8　6	(10)7 -5	(14)6—4	(18)5—3	(22)4—2	(26)3—8

续表

第一轮	第二轮	第三轮	第四轮	第五轮	第六轮	第七轮
(3)3—6	(7)2—5	(11)8—4	(15)7—3	(19)6—2	(23)5—8	(27)4—7
(4)4—5	(8)3—4	(12)2—3	(16)8—2	(20)7—8	(24)6—7	(28)5—6

表 3-11 中,圆括号里的数字代表场次,圆括号旁为比赛球队的队序号。

②奇数支球队赛程安排。相比于偶数支队伍,奇数支队伍的比赛要想给出一个尽可能公平的赛程,明显复杂多了。以 $n=9$ 为例进行介绍,相隔场次数的上限为 $r=3$,记 9 支球队为 $1,2,\cdots,9$,共 36 场比赛。赛程编制如下:

画一 4×9 的表格,如表 3-12 第 i 行第 j 列的格子记作(i,j),在每格左侧先按行依次填 1,3,5,7(第 1 行 1 个 1,第 2 行 3 个 3,第 3 行 5 个 5,第 4 行 7 个 7),后按行依次填 8,6,4,2,构成每场比赛的第 1 支队。

表 3-12 9 支球队参赛的比赛顺序的初始化

	1	2	3	4	5	6	7	8	9
1	1	8	8	8	8	8	8	8	8
2	3	3	3	6	6	6	6	6	6
3	5	5	5	5	5	4	4	4	4
4	7	7	7	7	7	7	7	2	2

在格的右侧沿各对角线填 1,3,5,7,如表 3-13 自(2,2)至(4,4)右侧填入 1,跳过一列再自(1,6)至(4,9)右侧填入 1,使 1 的总数(包括格子左侧的)为 8;自(3,4)至(4,5)右侧填入 3,跳过一列再自(1,7)至(3,9)右侧填入 3,使 3 的总数(包括格子左侧的)为 8;然后 5 和 7 类似;得表 3-13。

表 3-13 9 支球队参赛的比赛顺序的第一次轮转

	1	2	3	4	5	6	7	8	9
1	1	8	8	8	8	8—1	8—3	8—5	8—7
2	3	3—1	3	6	6	6	6—1	6—3	6—5
3	5	5	5—1	5—3	5	4	4	4—1	4—3
4	7	7	7	7—1	7—3	7—5	7	2	2—1

在格的右侧沿各对角线填 2,4,6,方法与上类似。最后在未满的 8 个格中填 9,得到表 3-14。按照表 3-14 先列后行的顺序排列得到赛程,即第 1 场 1 对 9,第 2 场 3 对 2,…,第 36 场 2 对 1。

表 3-14　9 支球队参赛的比赛顺序

	1	2	3	4	5	6	7	8	9
1	1—9	8—9	8—6	8—4	8—2	8—1	8—3	8—5	8—7
2	3—2	3—1	3—9	6—9	6—4	6—2	6—1	6—3	6—5
3	5—4	5—2	5—1	5—3	5—3	4—9	4—2	4—1	4—3
4	7—6	7—4	7—2	7—1	7—3	7—5	7—9	2—9	2—1

可以证明，如此编制出的赛程，其中每支队伍两场最大间隔达到最小，因此是最公平的赛程。

思考题

1. 铁饼投掷。如果说铅球因为其实心铅质球体，其投掷距离主要与出手时的速度、角度和高度有关，受空气阻力、风速影响较小。那么扁平的铁饼和细长的标枪，会明显受空气阻力、风速影响。早在公元前 708 年第 18 届古代奥运会中，掷铁饼就被列为 5 项竞技之一。如今铁饼是一个有金属包边的木质圆盘，中心的厚度较四周的厚度大，横截面类似于一个双凸的对称菱形翼型。投掷时，运动员被限制在直径 2.5 m 的圆形区域内，掷出的铁饼必须落在 40°的扇形区域内才有效。男子铁饼的规格：质量 2.005～2.025 kg，直径 21.8～22.1 cm；女子铁饼的规格：质量 1.005～1.025 kg，直径 18～18.2 cm。铁饼飞行的轨迹是重力和空气动力共同作用的结果，而这一结果又与铁饼出手时的角度、速度及自转角速度等初始条件密切相关。请建立 1 个铁饼投掷的数学模型。（提示：参阅陈永丽、朱克勤《铁饼飞行的轨迹和气动力》一文）

2. 请查阅陆长青的《第 29 届奥运会女篮 8 强队运动员年龄和身体形态特征的比较》一文，与本章男篮数据进行性别对比解读。

3. 2012 年伦敦奥运会中，来自浙江省 20 岁的孙杨获得了 400 m 和 1500 m 自由泳的冠军以及 200 m 的亚军。除了天然的身体条件和高超的划水技巧之外，他转身技巧的成熟应用也为比赛加了不少分，游泳的转身技巧有好几种，他用的是前滚翻转身，是大多数专业运动员自由泳比赛中都会用到的一种方法，也是所有游泳转身中最快的一种技术。请参考王丽等人的论文《自由泳转身阶段的生物力学研究》和参考文献中提到的资料，研究分析孙杨进一步提高游泳成绩的可能性。

4. 美国职业篮球联赛 NBA 赛程安排问题。NBA 共有 30 支球队，西部联盟、东部联盟各 15 支，大致按照地理位置，西部分西南、西北和太平洋 3 个区，东部分东南、中部和大西洋 3 个区，每区 5 支球队。NBA 正式赛季于每年 11

月的第一个星期的星期二开始，分为常规赛和季后赛两部分。常规赛为循环赛制，每支球队都要完成82场比赛，常规赛到次年的4月底结束，东西部联盟的前八名，包括各个赛区的冠军，将有资格进入接下来进行的季后赛。季后赛采用七战四胜赛制，共分四轮；季后赛的最后一轮也称为总决赛，由两个联盟的冠军争夺NBA的最高荣誉——总冠军。其中季后赛前三轮的赛程是采用2—2—1—1—1赛制，总决赛是采用2—3—2赛制（其中常规赛战绩好的球队有主场优势）。对于NBA这样庞大的赛事，编制一个完整的、对各球队尽可能公平的赛程是一件非常复杂的事情，赛程的安排对球队实力的发挥和战绩有一定的影响，从报刊上可以经常看到球员、教练和媒体对赛程的抱怨或评论。请针对NBA联盟公布的2012—2013赛程，设计评判标准，评价赛程安排的合理性，并指出有利和不利球队的名单。

第 4 章
真 伪 鉴 别

红楼梦“太虚幻境”中有一副对联，道是：“假作真时真亦假，无为有处有还无。”真真假假，假假真真；古往今来，有多少人物、多少事件让人无法看清真相；即便是白纸黑字的号称最正统的“史书”，又有多少水分？谁是《红楼梦》的作者？岳飞是《满江红》的词作者吗？莎士比亚真的是一个人写了那么多作品吗？伽利略在比萨斜塔做过自由落体运动实验吗？韩寒作品有“代笔”“水军”“包装”吗？这些广受争议的事件，也许有水落石出的一天，也许成为永远的谜。本章介绍用统计方法证伪文学作品以及用放射物衰变模型鉴定油画和文物。

4.1 《红楼梦》前 80 回和后 40 回的作者之谜

4.1.1 《红楼梦》简介

中国古代四大名著——《三国演义》《水浒传》《西游记》和《红楼梦》，均属于章回体长篇小说。四大名著是中国乃至全人类共同拥有的宝贵文化遗产，在整个华人世界中有着深远的影响。它们是在流传过程中广大读者研究、解读、衡量后评定出来的代表作品；承载了许多文化精华，对后人的思维智慧、为人处世方式产生了潜移默化的影响，对以后社会思潮、社会风气、政治生态的变化有着深远的影响。

《红楼梦》是一部有高度思想性和艺术性的伟大作品，代表古典小说艺术的最高成就，也是中国古代文学的典范之作。它以荣国府的日常生活为中心，以宝玉、黛玉、宝钗的爱情婚姻悲剧及大观园中的琐事为主线，以金陵贵族名门贾、史、王、薛四大家族由鼎盛走向衰亡的历史为暗线，展现了穷途末路的封建社会终将走向灭亡的必然趋势。并以其曲折隐晦的表现手法、凄凉深切的情感格调、强烈高远的思想底蕴，在我国古代民俗、封建制度、社会图景、建筑金石等各领域皆有不可替代的研究价值，达到我国古典小说的高峰，被誉为“我国封建社会的百科全书”。《红楼梦》到底属于什么样的书？历来众说纷纭。鲁迅将其定义为“人情小说”。脂砚斋《凡例》评：此书只是着意于

闺中，故叙闺中之事切，略涉于外事者则简。王国维《红楼梦评论》：《红楼梦》一书与一切喜剧相反，彻头彻尾之悲剧也。胡适《红楼梦考证》：《红楼梦》这部书是曹雪芹的自叙传。蔡元培《红楼梦索隐》：揭清之失，悼明之亡。

《红楼梦》在初期通过名为《石头记》的手抄本形式流传开来，自乾隆年间始，关于《红楼梦》的续作纷纷出炉。200多年来，人们不断地阅读它、谈论它、研究它，并衍生出一门国际性的学问——“红学”。红学研究不仅可以加深人们对《红楼梦》这部辉煌巨著的认识和理解，而且对于古典文学研究的现代化探索也有借鉴意义。《红楼梦》的著作权问题是红学研究的一个重要问题，作者是谁，一直是人们争论的焦点。曾经不少学者认为《红楼梦》的前80回是曹雪芹所写，后40回为高鹗所续。但也有人提出质疑，最极端的是，甚至有人说《红楼梦》的真正作者是李鼐，苏州织造李煦之子，而不是曹雪芹。《红楼梦》的作者究竟是谁？人们曾从各个方面进行考证、分析，令人感兴趣的是，一些专家用统计的方法来研究《红楼梦》的作者问题，取得了似乎最客观、最让人信服的成果。

关于《红楼梦》版本的常见说法是，《红楼梦》共120回，前80回由曹雪芹作，后40回由无名氏续，程伟元、高鹗整理；曹作八十回在撰写、修改过程中就以抄本的方式流传。乾隆五十年(1785年)，程伟元将前80回及后40回续稿以活字排印，从此120回本流行。但前80回的文字曾有改动。它的原名有《石头记》《风月宝鉴》《金陵十二钗》《情僧录》《大观琐录》《金玉缘》《情界铨》等，梦觉主人序本正式题为《红楼梦》。《红楼梦》有十几个版本，分程高本、脂本两大系统，通行本为程甲本。脂本是20世纪新发现的“古本”，通名《脂砚斋重评石头记》，其中最受关注的是甲戌本、己卯本、庚辰本。2010年人民文学出版社新版《红楼梦》署名“曹雪芹著、无名氏续”，抛弃了原来的“高鹗续书说”。

4.1.2 《红楼梦》统计分析

一般地说，作家写作时某处的遣词用字由于其写作习惯与文风的影响，往往表现出一定的规律性。如果整本小说是出自一人的话，那么不管是书的哪一部分，其遣词用字的写作手法都应该是相像的。这是文学作品真伪辨别的最大前提。否则的话，要么是原著者因某种特殊原因(如重大人生变故，改变了原有的人生观)改变了写作风格，要么是另一人续写后半部。纵观古今中外的作家们，随着年龄的增长、人生阅历的丰富，创作风格一般都会发生变化，早期作品与晚期作品几乎在许多地方存在明显的差异。但在同一部作品中，通常一名作家的创作思路有着明确的连贯性，某些词语或某种表现手法

会特别频繁地出现，就像人们说话时会有习惯用语或口头禅一样。这些也就促使人们借用统计方法来研究辨别作品真伪。

1980 年 6 月，美国威斯康星大学陈炳藻教授在首届国际《红楼梦》研讨会上宣读了一篇《从词汇上的统计论〈红楼梦〉的作者问题》的文章，引起了国际红学界的注意和兴趣。1986 年，陈炳藻教授公开发表了《电脑在文学上的应用:〈红楼梦〉与〈儿女英雄传〉两书作者用词的比较》一文；之后又出版了《电脑红学:论〈红楼梦〉作者》的专著。陈炳藻教授通过利用电脑对《红楼梦》前 80 回和后 40 回的用字进行了测定，并从数理统计学的观点出发，探讨《红楼梦》前后用字的相关程度。他将《红楼梦》120 回按顺序编成 3 组，每组 40 回。并将《儿女英雄传》作为第 4 组进行比较研究，从每组中任取 8 万字，分别挑出名词、动词、形容词、副词、虚词这 5 种词，借助统计上的相关分析计算各组相关程度。结果发现，《红楼梦》前 80 回与后 40 回所用的词汇正相关程度达 78.57%，而《红楼梦》与《儿女英雄传》所用词的正相关程度是 32.14%。由此推断得出前 80 回与后 40 回的作者均为曹雪芹一人的结论。

然而，复旦大学数学系李贤平副教授却给出了相反的回答。1987 年李贤平带领他的学生将 120 回看成是 120 个样本，然后确定与情节无关的 47 个虚词(之、其、或、亦……，呀、吗、咧、罢……，的、着、是、在……，可、便、就、但……)作为变量(同一情节描述用词差不多)，统计出每一回里变量出现的次数作为数据，用多元统计中的聚类分析法进行合并，果然将 120 回分成两类，即前 80 回为一类，后 40 回为一类，很形象地证实了不是出自同一人的手笔。之后他们又进一步分析了前 80 回是否为曹雪芹所写这一问题，根据曹雪芹的另一本著作做了类似计算，结果证实了用词手法完全相同，断定为曹雪芹一人的手笔。类似的论证还推翻了后 40 回是高鹗一人所写的传统认识。

虽然应用统计技术来研究《红楼梦》这样伟大的文学作品，还不能成为红学研究的主流，但已经引起人们的广泛关注，可以说已成为文学研究的一个不可或缺的辅助工具。同样运用统计方法，得到的结论会完全不同，这又是为什么呢？相比较而言，李贤平的研究统计分组更详细，剔除情景用词的设计更合理，使用的统计方法也更高级，得到的结论可信度可能会更高一些。2010 年浙江大学汪维辉在前人研究的基础上，利用电子本进行穷尽性检索，举出更多的例子来证明前 80 回和后 40 回之间存在的词汇差异。他论列了 100 多个词语，大体分为四类:同词异义、同义异词、同词而使用频率或用法不同、仅见于前 80 回或后 40 回的词语。结论是:前 80 回和后 40 回的词汇存在明显差异，不可能是一人所作。

中国台湾林试所森林生物系主任潘富俊著作《红楼梦植物图鉴》(上海书

店出版社，2005年)，将红楼梦中提及的200多种植物作为统计研究对象，先将这部120回的小说等分为三个部分，发现第一个40回平均每回出现11.2种植物，中间40回平均每回出现10.7种，最后40回平均数只有3.8种。潘富俊还说，后40回中只有6成提到茶，且只有一种茶——龙井茶；前80回逾92%提到茶，且有9种茶。这个统计结果表明，《红楼梦》的作者可能不是出自同一人。

有人发现，在《红楼梦》的前80回里，丫环、佣人、老妈子等下人都自称为"小的"，到了后40回，则一般都自称为"奴才"，这露出了一个"马脚"，据此得出《红楼梦》的前80回与后40回的作者可能不是同一个人。也有一位学者选出"儿、在、事、的、著"5字，对前80回与后40回分别作了统计、比较，发现这5字出现的频率相当稳定，虽出自于两人之手，但后40回仿前80回使用的口语是"有过之"而"无不及"。

与《红楼梦》类似，用统计的观点考证文学著作的作者还有一个更为典型的例子。苏联有一部文学名著《静静的顿河》，在出版时署名作者为肖洛霍夫，出版后不久，就有人说这本书是从哥萨克作家克留柯夫那里抄袭而来的。为了弄清谁是真正的作者，一些学者用统计的方法进行考证，从句子的平均长度、词的选用、结构分析、用词频率等作了统计方法分析，最后确认该书作者确为肖洛霍夫。

运用统计方法分析文学作品的最关键的地方是，作者创作是否一定保持高度连贯性、能否出现开小差现象？譬如，开始时非常注重植物(包括茶)在其中的烘托作用以显示家族的荣华富贵，后来随着主人公家境衰落而无需过多植物(包括茶)的描述，也许是故事情节演绎时影响作者心理状态，忘记了植物(包括茶)的参与。

4.2 统计检验

4.2.1 假设检验

俗话说"大千世界，无奇不有"，"不怕不识货，只怕货比货"。我们生活在"五光十色"的世界里，生物多样性、文化多样性、社会分工细化等构成了这一大千世界，"异同共存"是大自然和人类社会共同遵循的基本原则。如何界定衡量什么叫"相异"、什么叫"相同"？假如事物A与事物B有多处不一样，被认为相异。但实际上两者差异似乎不大，说不定这点差异是因为观测误差所带来的。又譬如，奶粉包装袋上注明净含450 g，人们自然要求所有这些袋装奶粉都应该刚好450 g，奶粉生产商做得到吗？要想做到毫厘无差，必然是先

计量包装袋重量，再计量包装了奶粉的重量，如果有差异得重新操作。如此操作，每天一条生产线一个工人恐怕只能包装数量可怜的奶粉，这家生产商除了关门别无选择。如果允许少量的误差，这样可大大提高生产效率，降低成本，使普通百姓消费得起。实际上如果装上 448 g，消费者根本感觉不出，而且 2 g 的差异根本不会影响到消费。

现在的问题是，到底相差多少才算被容许？如果说某袋只装了 350 g，整整少了 100 g，相当于标准的 100/450=2/9，这对于几乎所有消费者来说都是无法忍受的。显然，这里主要是要界定多少差异可认为是随机误差造成的，或者认为一般随机误差不应该带来如此大差异因而考虑操作不当或故意偷工减料造成的。

假设检验是数理统计学中根据一定假设条件由样本推断总体的一种方法。在假设检验问题中，由于作出判断的依据是样本，随机因素的影响难免会产生判断错误，我们把可能犯的错误分为两类：当原假设 H_0 为真时，根据样本观测值却做出了拒绝 H_0 的判断，我们称之为犯第一类错误；当 H_0 为假时，根据样本观测值却做出了接受 H_0 的判断，我们称之为犯第二类错误。犯两类错误的概率分别记为：

$$P(\text{拒绝 } H_0 \mid H_0 \text{ 为真})=\alpha, P(\text{接受 } H_0 \mid H_0 \text{ 为假})=\beta$$

在样本容量一定时，两类错误具有跷跷板现象，若要减小犯一类错误的概率，则犯另一类错误的概率往往会增大。要想同时减小犯两类错误的概率，只有增大样本容量。当样本容量一定时，一般我们总是先考虑控制犯第一类错误的概率 α，一般取 0.1、0.05、0.01 等值，而对犯第二类错误的概率不予考虑。这类检验问题称为显著性检验问题。

形如 $H_0:\mu=\mu_0, H_1:\mu\neq\mu_0$ 中，备择假设 H_1 表示 μ 可能大于 μ_0，也可能小于 μ_0，称为双边备择假设，这类假设检验的拒绝域位于两侧，我们称之为双侧（双边）假设检验。如果将需要检验的问题写成以下的形式：

$$H_0:\mu\leqslant\mu_0, H_1:\mu>\mu_0, \text{或者 } H_0:\mu\geqslant\mu_0, H_1:\mu<\mu_0$$

这类假设检验的拒绝域位于一侧，我们称之为单侧（单边）假设检验。在处理具体问题时，通常把着重考察并且便于处理和等号成立的假设作为原假设。有关假设检验的详细内容几乎可以在任何一本《概率论与数理统计》的书中找到。

4.2.2 韦博成的工作

本小节介绍东南大学著名统计学家韦博成的研究成果“《红楼梦》前 80 回与后 40 回某些文风差异的统计分析”的主要内容。

原假设 H_0:"前80回与后40回对于'饮食描写'的关注程度相同";

对立假设 H_1:"前80回对于'饮食描写'的关注程度大于后40回对于'饮食描写'的关注程度";

选择《红楼梦》中着力描写的5个情景指标,即花卉、树木、饮食、医药与诗词,统计出它们在前80回与后40回中出现的频数,如表4-1。

表4-1 前80回与后40回各情景指标出现的频数

	1—40回	41—80回	1—80回	81—120回
花卉	15	16	31	7
树木	13	14	27	7
饮食	17	17	34	8
医药	13	13	26	8
诗词	22	14	36	12

《红楼梦》前80回与后40回在文风上的差异分析可化为数理统计学的问题。以饮食的描写为例。假设

(1)$X_1 \sim B(n_1, p_1)$表示前80回的二项分布,其中 $n_1=80$,X_1 表示前80回中涉及饮食的回数;其实测值为 $X_1=34$,p_1 表示前80回中每回涉及饮食的概率。

(2)$X_2 \sim B(n_2, p_2)$表示后40回的二项分布,其中 $n_2=40$,X_2 表示后40回中涉及饮食的回数;其实测值为 $X_2=8$,p_2 表示后40回中每回涉及饮食的概率。

等价性检验问题为:

$$H_0: p_1 = p_2, \longleftrightarrow H_1: p_1 > p_2$$

否定原假设就意味着,在一定显著性水平下"前80回对于'饮食描写'的关注程度大于后40回对于'饮食描写'的关注程度"。

先粗略计算一下,前80回大约平均每回出现34/80=0.425次饮食描写,后40回大约平均每回出现8/40=0.20次饮食描写。0.425与0.20之间差异算大还是算小?如果后40回平均每回出现0.40次描写饮食,即0.425与0.40之间,我们可能就会认为彼此之间差异很小,几乎可以认为近似相等。但是现在是0.425与0.20,似乎有些大,那么应该如何说明?

在差异性检验时通常总是假设没差异,没差异就意味着两者相等,那么都等于多少呢?直观地看,人们容易想到用某种平均数表示,一般建议取$(X_1+X_2)/(n_1+n_2)$,此处$(34+8)/(80+40)=0.35$。现在的问题变成了,在平均每回0.35次饮食描写的假设下,出现前80回34次和后40回8次的结果

是否算正常？在数理统计的假设检验里有两种做法：Fisher 精确条件检验和渐近正态检验。其中渐近正态检验的检验统计量为

$$u=\frac{X_1/n_1-X_2/n_2}{\sqrt{\hat{p}(1-\hat{p})(n_1^{-1}+n_2^{-1})}}\sim N(0,1)$$

其中 $\hat{p}=(X_1+X_2)/(n_1+n_2)$。精确条件检验公式比较复杂，略。表 4-2 给出了各个情形指标的检验结果。

表 4-2　前 80 回与后 40 回各情景指标检验结果

	u 检验值	u 检验的 p-值	精确条件检验的 p-值
饮食	2.4360	0.0074	0.0114
花卉	2.3590	0.0092	0.0140
树木	1.8622	0.0313	0.0473
诗词	1.5811	0.0569	0.0824
医药	1.4325	0.0760	0.1105

u 检验是一种近似的检验，表 4-2 显示它的 p-值与精确条件检验的 p-值相近，但略偏小。在统计检验中，p-值表示在原假设成立的前提下，出现实际观测结果数据以外的概率，或者当作出拒绝原假设时可能判断错误的概率。比较饮食描写次数精确条件检验的 p-值为 0.0473，即在假设两者都是按每回出现 0.35 次饮食描写下，出现前 80 回 34 次和后 40 回 8 次的结果至多是 0.0473。0.0473＝4.73％，即 100 次试验中至多有 5 次。在概率统计中，有一个非常有名的小概率不发生原理，一个理论概率很小的事件在一次试验中被认为是几乎不发生的。其中小概率的具体值通常约定为 0.01、0.05、0.10，特别是 0.05。饮食描写比较的 p-值等于 0.0114 很小，完全可以认为前 80 回与后 40 回饮食指标存在明显差异。类似地，花卉和树木的描写差异显著，而诗歌与医药描写似乎有一点差异。

4.3　莎士比亚新诗真伪之鉴定

4.3.1　莎士比亚简介

威廉·莎士比亚享誉世界，其不朽作品广为流传。然而，莎士比亚真有其人吗？世界文化巨人威廉·莎士比亚是欧洲文艺复兴时期最伟大的戏剧家和杰出的诗人，他出生于英国埃文河畔斯特拉特福镇的一个商人市民家庭。21 岁时离家外出谋生，当过剧场的杂役、演员，继而成为一个剧作家。莎士比亚一生中创作了 37 部戏剧、154 首十四行诗和 2 首长诗。除了 2 首长诗

是他生前发表的以外，莎氏的全部作品都是在他死后由别人收集成书的。一直存在着这么一个问题，以演员威廉·莎士比亚的名字发表的那些精彩伟大的作品，究竟是他本人写的还是别人创作后用假托的笔名发表的？莎士比亚是众人皆知的伟大剧作家，但他的身世却有许多不为世人所知之处，他未曾留下只言片语。有关莎士比亚生平的材料奇缺。在莎士比亚女婿霍尔医生的日记中，也找不到其岳父是著名剧作家的任何说明。不可思议的是，当时没有一个人能明确说明哪些作品是演员莎士比亚创作的。莎士比亚去世也没有引起任何人的重视，没有一个人按照当时的习俗为他写一首哀诗。因此，即使像拜伦和狄更斯这样的大作家也怀疑莎士比亚是否写过那些杰作，狄更斯还曾经表示一定要揭开“莎士比亚真伪之谜”。当然这一谜底是永远不可能解开的。

4.3.2 莎士比亚新诗

1985年11月14日，英国及美国的报纸及杂志争相报道，研究莎士比亚的美国学者 Gary Taylor 在英国牛津大学的 Bodleian 图书馆的一本于1755年收藏的书中，找到一首很可能是莎士比亚的抒情诗，该诗的第二抄本稍后在美国耶鲁大学的图书馆出现，我们姑且先称这首诗为泰勒诗。如果能证实泰勒诗确为莎士比亚所作，那么这将是17世纪以来莎士比亚作品最重要的一次发现。英美学者为了这首诗之真伪争论不休，大打笔战。

只是不少专家认为这首泰勒诗不论在用字遣词与韵味风格上，都迥异于莎士比亚的其他作品。有趣的是，统计学者也介入这场纷争。1986年1月24日出版的 Science 杂志，刊登了一篇文章——《莎士比亚的新诗——向统计学礼赞》，介绍著名的统计学者——美国斯坦福大学的 Efron 教授及芝加哥大学的 Thisted 教授，如何以统计的方法鉴定这首新出土的泰勒诗是否为莎士比亚所作。

4.3.3 统计学家 Efron 和 Thisted 的想法

早在1976年，Efron 和 Thisted 便把莎士比亚作品中所用的字做了一番统计分析。他们想回答若发现了一个新的作品，如何经由统计分析其中所用的字以确定此作品是否为莎士比亚所作。他们当初做此分析只是为了好玩，没想到10年后真派上了用场。

这并非统计学家第一次协助解决文学上的问题。由于统计分析如此具有说服力，往往能使文学上长期的争论迅速地平息。例如，哈佛大学的 Mosteller 及 Wallace 利用统计方法，判定 *The Federalist Papers* 的作者为

James Madison，而非 Alexander Hamiliton。在 1950 年，著名的英国统计学者 David Cox 及文学家 L. Brandwood 也曾利用统计方法，解决争论长达 1000 年的关于柏拉图众多作品的先后顺序。

虽然利用统计方法来解答文学上的问题之想法并非创新，Efron 与 Thisted 所采用的特别方法却不曾在这方面用过。此方法可追溯至 1940 年，那时生物学家 Williams 向英国统计学家费希尔(R. Fisher)提出一个似乎不可能回答的问题。Williams 曾前往马来西亚采集蝴蝶，他把自己共见过几种蝴蝶、每种各见过的次数都告诉 Fisher。Williams 想知道，在马来西亚的蝴蝶中他没见过的究竟有多少。

一般人会觉得此问题毫无头绪，不可能得到解答。但统计学家却有办法估计尚有几种蝴蝶还未被捕捉到。只要假设蝴蝶是依照每一种之只数的比例，随机地被捕捉。而这只要假设每一件事(包括蝴蝶的分布、捕捉的技术等)随时且很均匀，不会先是把某一种蝴蝶捕捉殆尽，之后再大量捕捉另一种。若有某一种蝴蝶尚未被捕捉到，那纯粹只是运气的关系，而非该种蝴蝶特别会躲藏。Fisher 所用方法之细节，由于技术性强，此处无法细说。

Efron 与 Thisted 介入文学是很偶然的。有一次他们聆听加州大学 Santa Barbara 分校的 Gani 的演讲，Gani 的目的是要分析莎士比亚作品的结构。在演讲中他提到德国 Münster 的 Westfälische Wilhelms-Universität 的 Marvin Spevack，已将莎士比亚的所有作品输入计算机中，并已计算出莎士比亚所用过的全部字数及每一个字使用过的次数。听完演讲后，Efron 与 Thisted 决定把 Fisher 所用的方法拿来分析莎士比亚的作品，他们的研究报告发表在统计学中极具权威性的学术期刊 Biometrika 上。

这一篇文章的题目为 *Estimating the Number of Unseen Species: How Many Words Did Shakespeare Know?* 如前所述，在生态学中往往会估计某生物尚未见到的种数。而在该文中，尚未见到的“种”，却是莎士比亚知道但不曾用过的字。

莎士比亚全部作品(以下简称总作品)之总字数为 884647，其中有 14376 个相异字只出现 1 次，4343 个相异字只出现 2 次。表 4-3 列出出现 100 次的字数。令 n_x 表示出现 x 次的字数。当 $x=43$，即列 40+，行 3，其对应的 $n_x=30$，表示出现 43 次的字共有 30 个，余类推。

表 4-3　莎士比亚总作品中字出现之频数

x	0	1	2	3	4	5	6	7	8	9	小计
0+	14376	4343	2292	1463	1043	837	638	519	430	364	26305
10+	305	259	242	223	187	181	179	130	127	128	1961
20+	104	105	99	112	93	74	83	76	72	63	881
30+	73	47	56	59	53	45	34	49	45	52	513
40+	49	41	30	35	37	21	41	30	28	19	331
50+	25	19	28	27	31	19	19	22	23	14	227
60+	30	19	21	18	15	10	15	14	11	16	169
70+	13	12	10	16	18	11	8	15	12	7	122
80+	13	12	11	8	10	11	7	12	9	8	101
90+	4	7	6	7	10	10	15	7	7	5	78

在总作品中，莎士比亚共用了 31534 个不同的字，即

$$n_1+n_2+n_3+\cdots+n_x+\cdots=31534$$

必须注意的是像“girl”与“girls”乃视为不同的字。由表 4-3，出现次数不超过 100 的有 30688 个字，因此有 846 个字出现的次数超过 100。

Efron 与 Thisted 分别依据 Fisher 等人所用的参数模式及 Good 和 Toulmin 所用的非参数模式来估计莎士比亚尚认识多少字，所得到的估计值都差不多是 11460，且标准差小于 150。

事实上，在 Efron 和 Thisted 所著该文中，他们想回答更一般的问题。他们写着“假设有另一很大量的莎士比亚的作品被发现，譬如说共有 884647 个字，则除了那 31534 个不同的字外，预期可找到几个新字?”莎士比亚共认识几个字就对应于∞的情况。由于自 17 世纪之后，便没有莎士比亚的新作品出现，所以 Efron 与 Thisted 从未想过会有机会真正用莎士比亚的作品来检验他们的理论。他们做此研究的动机纯粹是觉得有趣。甚至当他从新闻上获知这首泰勒诗出现，且有可能是莎士比亚所作时，他一时并没想到他其实曾与 Thisted 做过这方面的研究。直到 Thisted 提醒他，他才想到 10 年前所做的工作。尘封已久的莎士比亚研究又浮现在眼前。

4.3.4　Efron 和 Thisted 的检验结果

这首泰勒诗与总作品相比，相当短，总共只有 429 个字。Efron 与 Thisted 基于罕用字出现的频率，发展出一种在统计上算是简单的检定法。

有一些字是大家常用的，如“a，is，the”，在每一篇文章中可能都出现许多次。但对罕用字，每个作者使用的习惯可能就不同了。Efron 与 Thisted 指出，在总作品中罕用字的使用非常普遍，在全部使用的 31534 个相异字中，有接近 2/3 的比例(共 21011 个)，只用了不超过 3 次(见表 4-3)。

这首泰勒诗虽只有 429 个字，但其中包含 258 个相异字。令 m_x 表示在泰勒诗中所使用的字，在总作品中出现 x 次的相异字数。表 4-4 列出 $x=0$ 至 99 的统计。例如，泰勒诗中有 7 个相异字在总作品中只出现过 1 次，即 $m_1=7$，有 5 个相异字在总作品中出现过 23 次，即 $m_{23}=5$，余类推。表 4-4 中包含 118 个相异字，泰勒诗中另有 140 个相异字在总作品中出现 100 次以上。其中特别值得注意的是 $m_0=9$，此为泰勒诗中不曾在总作品中出现的相异字数。这就是 Efron 与 Thisted 所要估计的量，记作 $\Delta(t)$。

表 4-4　泰勒诗在总作品中出现 x 次的相异字 m_x

x	0	1	2	3	4	5	6	7	8	9	小计
0+	9	7	5	4	4	2	4	0	2	3	40
10+	1	0	3	0	1	1	1	2	1	0	10
20+	2	2	1	5	3	1	0	2	2	3	21
30+	4	1	1	1	2	1	0	0	3	3	16
40+	1	2	0	0	2	1	1	2	1	1	11
50+	0	1	1	1	1	0	0	1	0	2	7
60+	0	1	0	0	1	1	0	0	1	0	4
70+	0	0	1	0	0	1	0	0	1	1	4
80+	0	0	1	1	0	0	0	0	0	0	2
90+	0	0	0	1	0	1	1	0	0	0	3

Efron 与 Thisted 将他们的结果整理成论文，题目就叫 *Did Shakespeare Write a Newly Discovered Poem*？表 4-5 为他们求出的若泰勒诗真为莎士比亚所作，表 4-4 中 m_x 之期望值的估计值。由表 4-5 知，若此泰勒诗确为莎士比亚所作，则 Efron 与 Thisted 估计其中含有 $\Delta(t)=6.97$ 个在总作品中未曾出现的相异字。若再考量标准差，则此首新诗约包含 6.97 ± 2.64 个新字，即新字的数目约介于 4.33 与 9.61 之间。实际的新字有 9 个(即表 4-4 中的 m_0)，的确在 4.33 与 9.61 之间。估计曾出现一次的字有 4.21 ± 2.05 个，实际则为 7，估计曾出现两次的字有 3.33 ± 1.83 个，实际则为 5。Efron 与 Thisted 一直分析到曾出现 100 次的字，其吻合程度皆相当惊人，可以说通过

严格的统计鉴定，看起来这首诗确是莎士比亚所写，或者用比较保守的统计术语来说：没有足够的证据可推翻此诗为莎士比亚所做之假设。

表 4-5　表 4-4 中 m_x 之期望值的估计值 ν_x

x	0	1	2	3	4	5	6	7	8	9
0+	6. 97	4. 21	3. 33	2. 84	2. 53	2. 43	2. 16	2. 01	1. 87	1. 76
10+	1. 62	1. 50	1. 52	1. 51	1. 36	1. 38	1. 33	1. 28	1. 25	1. 22
20+	1. 18	1. 16	1. 13	1. 11	1. 09	1. 06	1. 04	1. 02	1. 00	0. 98
30+	0. 96	0. 94	0. 93	0. 91	0. 90	0. 88	0. 86	0. 85	0. 83	0. 82
40+	0. 80	0. 79	0. 77	0. 76	0. 75	0. 74	0. 73	0. 72	0. 70	0. 69
50+	0. 68	0. 67	0. 66	0. 65	0. 64	0. 63	0. 62	0. 61	0. 60	0. 59
60+	0. 58	0. 57	0. 56	0. 55	0. 54	0. 53	0. 52	0. 51	0. 50	0. 50
70+	0. 49	0. 48	0. 48	0. 47	0. 47	0. 46	0. 45	0. 45	0. 44	0. 44
80+	0. 43	0. 42	0. 42	0. 41	0. 41	0. 40	0. 39	0. 39	0. 38	0. 38
90+	0. 37	0. 36	0. 36	0. 35	0. 35	0. 34	0. 34	0. 33	0. 32	0. 32

传奇统计学者 Persi Diaconis 熟悉 Efron 及 Thisted 的分析。据他讲，他第一次读这首泰勒诗时，觉得这一点也不像莎士比亚的作品，他认为只要做一些数值分析，就可以证明诗中字所摆的位置完全是错的。但读了 Efron 及 Thisted 的分析后，他相信这首诗很可能是莎士比亚所作。

Efron 强调他们的分析并不能证明这首泰勒诗真是莎士比亚写的。但他说这首泰勒诗在罕用字的使用上，如此吻合莎士比亚的总作品，确实令人惊讶。Efron 及 Thisted 也对 John Donne、Christophor Marlowe 及 Ben Jonson 等 3 位约略与莎士比亚同时代的诗人，各取 1 首诗，再另取 4 首莎士比亚的诗，与这首泰勒诗做比较，列出 8 首诗用字在总作品中出现的频率。经过三种严密的统计鉴定，发现对前 3 首(非莎士比亚之作品)罕用字出现次数之实际值与预测值(假设其为莎士比亚所作)并不吻合。而虽然挑选的 4 首莎士比亚的诗偶尔有不吻合处，总的来说是可接受的。本来仅是师生的游戏之作，说不定当初还被“有识之士”视为无稽之谈，10 年后竟令那些向来可能不常接触统计的文学学者折服。不过与其将此视为无心插柳，不如相信统计学的无所不在。

4.4 油画造假与鉴定

4.4.1 人体蒋碧薇女士

2010 年 6 月,北京九歌国际拍卖有限公司在春季拍卖中,以 7280 万元人民币的价格,成功拍出了名为《人体蒋碧薇女士》的"徐悲鸿油画"(图 4-1)。该油画的拍卖信息被广为散布,同时配发的还有"徐悲鸿长子徐伯阳与这幅画的合影"以及徐伯阳出示的"背书",背书内容为:"此幅油画(人体)确系先父徐悲鸿的真迹,先父早期作品,为母亲保留之遗作。徐伯阳 2007 年 9 月 29 日"。



图 4-1 名为《人体蒋碧薇女士》的"徐悲鸿油画"



图 4-2 徐悲鸿 1941 年作的蒋碧薇肖像作品

然而,2011 年 9 月的一天,中央美术学院首届研修班的部分学生联名声称,这幅油画是当年他们的习作。7280 万,抢到的是旷世名作还是废纸一张?中央美术学院油画系研修班第一届的 10 位同学以《这是徐悲鸿的作品,还是我们的习作?》为题,联名发出一封公开信,信中称,一幅售价 7000 多万的"徐悲鸿油画",其实只是当年这些中央美院孩子的习作。公开信发出当天,著名画家陈丹青立即发表评论,称《人体蒋碧薇女士》连"伪作"都算不上。陈丹青说:"用不着依据,我们都画了四五十年油画了,这个是一个简直不用去判断的事情,所以我比较惊讶,现在人最起码的比较都做不出来,这 20 年代和 80

年代的区别,江南小姐和北方丫头的区别已经基本看不出来了。”

据报道,2005年11月,俄罗斯著名画家鲍里斯·米哈伊洛维奇·库斯妥基耶夫的油画《宫女》被俄罗斯石油大亨维克托·斐克塞伯格拍得,2009年被鉴定为赝品;2006年,德国知名拍卖商蓝波次以破纪录的290万欧元价格售出坎彭东克的《马匹构成的红色画像》,随后被指为赝品;2008年,上海买家苏敏罗以253万拍到一幅吴冠中先生的《池塘》,在找吴老先生鉴定时,吴冠中亲笔写下“此画非我所作,系伪作”。2009年香港佳士得春拍中,一幅声称是吴冠中所作的油画《松树》,以158万港元成交,日后也被吴先生认定为伪作。2011年,故宫一群顶尖的鉴宝专家为了收取几十万元鉴定费而给一具由骗子自己制作的假金镂玉衣出具了一份价值24亿元的鉴定书,最终骗子以此诈骗了银行数亿元贷款并挥霍一空。

全世界收藏爱好者无以计数,由于巨大的利益驱使,据称古玩市场的赝品率高达90%以上,作假手段近年来层出不穷。

4.4.2 梅格伦造假

在第二次世界大战比利时解放以后,荷兰野战军保安机关开始搜捕纳粹同谋犯。他们从一家曾向纳粹德国出卖过艺术品的公司中发现线索,于1945年5月29日以通敌罪逮捕了三流画家范·梅格伦(H. A. Vanmeegren),此人曾将17世纪荷兰名画家扬·弗米尔(Jan Veermeer)的油画《捉奸》等卖给纳粹德国戈林的中间人。可是,范·梅格伦于同年7月12日在牢里宣称:他从未把《捉奸》卖给戈林,而且他还说,这一幅画和众所周知的油画《在埃牟斯的门徒》以及其他四幅冒充弗米尔的油画和两幅德胡斯(17世纪荷兰画家)的油画,都是他自己的作品,这件事在当时震惊了全世界,为了证明自己是一个伪造者,他在监狱里开始伪造弗米尔的油画《耶稣在门徒们中间》,当这项工作接近完成时,范·梅格伦获悉自己的通敌罪已被改为伪造罪,因此他拒绝将这幅画交出,以免留下罪证。

然而,事情并未到此结束,许多人还是不肯相信著名的《在埃牟斯的门徒》是范·梅格伦伪造的。事实上,在此之前这幅画已经被文物鉴定家认定为真迹,并以17万美元的高价被伦布兰特学会买下。专家小组对于怀疑者的回答是:由于范·梅格伦曾因他在艺术界中没有地位而十分懊恼,他下决心绘制《在埃牟斯的门徒》,来证明他高于三流画家。当创造出这样的杰作后,他的志气消退了。而且,当他看到这幅《在埃牟斯的门徒》多么容易卖掉以后,他在炮制后来的伪制品时就不太用心了。这种解释不能使怀疑者感到满意,他们要求完全科学地、确定地证明《在埃牟斯的门徒》的确是一个伪造品。

这一问题一直拖了 20 年，直到 1967 年，才被卡内基·梅隆大学(Carnegie Mellon University)的科学家们基本解决。

4.4.3 油画鉴定的杀手锏

(1)放射性与半衰期

测定油画和其他岩石类材料年龄的关键是 20 世纪初发现的放射性现象。在 20 世纪初，著名物理学家卢瑟夫发现，某些元素的原子是不稳定的，并且在已知的一段时间内，有一定比例的原子自然蜕变而形成新元素的原子。这一特征称为“放射性”，而这些元素称为放射性元素。卢瑟夫还发现，物质的放射性与所存在的物质的原子数成正比。

用 $N(t)$ 表示时间 t 时存在的原子数，则：

$$\mathrm{d}N(t)/\mathrm{d}t=-\lambda N(t)$$

其中 λ 是正的常数，称为该物质的衰变常数。

已知 t 在 t_0 时刻，该物质的原子数为 N_0，即 $N_0=N(t_0)$。求解前面的微分方程得 t 时刻该物质的原子数：

$$N(t)=N(t_0)\times\exp\{-\lambda(t-t_0)\}$$

设 $N(t)/N(t_0)=1/2$，$T=t-t_0=\ln 2/\lambda$。其中 T 为原子数从起初 t_0 时刻 $N(t_0)$ 衰变到只剩下一半时所耗费的时间，因此 T 也称为半衰期。许多物质的半衰期已被测定，如碳 14，5568 年；铀 238，45 亿年。

(2) 油画与放射性物质的相关知识

艺术家们应用白铅作为颜料之一，已达两千年以上。白铅中含有微量的放射铅 210，白铅是从铅矿中提炼出来的，而铅又属于铀系。

地壳里几乎所有的岩石中均含有微量的铀。一方面，铀系中的各种放射性物质均在不断衰减；而另一方面，铀又不断地衰减，补充着其后继元素。从而，各种放射性物质(除铀以外)在岩石中处于放射性平衡中。根据世界各地抽样测量的资料，地壳中的铀在铀系中所占平均重量比约为 27/1000000(一般含量极微)。各地采集的岩石中铀的含量差异很大，但从未发现含量高于 2%～3%的。

从铅矿中提炼铅时，铅 210 与铅 206 一起被作为铅留下，而其余物质则有 90%～95%被留在矿渣里，因而打破了原有的放射性平衡。

(3)问题简化约定

本问题建模是为了鉴定几幅不超过 300 年的古画，为了使模型尽可能简单，可作如下假设。

①由于镭的半衰期为 1600 年，经过 300 年左右，应用微分方程方法不难

计算出白铅中的镭至少还有原量的90%，故可以假定，每克白铅中的镭在每分钟里的分解数是一个常数。

②铅210的衰变为钋210和铅206：

铅210→(半衰期22年)→ 钋210 →(半衰期138天)→ 铅206

若画为真品，颜料应有300年左右或300年以上的历史，容易证明：每克白铅中钋210的分解数等于铅210的分解数(相差极微，已无法区别)。可用前者代替后者，因钋的半衰期较短，易于测量。

(4)建模

①记提炼白铅的时刻为 $t=0$，当时每克白铅中铅210的分子数为 y_0，由于提炼前岩石中的铀系是处于放射性平衡的，故铀与铅的单位时间分解数相同。由此容易推算出每克白铅中铅210每分钟分解数不能大于30000个，否则铀的含量将超过4%，而这是不可能的。因为：

若 $\lambda_U U_0=\lambda y_0\geqslant 30000$，则 $U_0\geqslant 30000\times 60\times 24\times 365/\lambda_U\approx 1.02\times 10^{20}$(个)，这些铀约重$[1.02\times 10^{20}/(6.02\times 10^{23})]\times 238\approx 0.04$(克)　　(1)

即每克白铅约含0.04克铀，含量为4%。以上确定了每克白铅中铅分解数的上界，若画上的铅分解数大于该值，说明画是赝品；但若是小于该值，不能断定画一定是真品。

②设 t 时刻1克白铅中铅210含量为 $y(t)$，而镭的单位时间分解数为 r(常数)，则 $y(t)$满足微分方程：

$$\mathrm{d}y/\mathrm{d}t=-\lambda y+r$$

由此解得：

$$y(t)=(r/\lambda)\times[1-\exp\{-\lambda(t-t_0)\}]+y(t_0)\times\exp\{-\lambda(t-t_0)\}$$

故：

$$\lambda y(t_0)=\lambda y(t)\times\exp\{\lambda(t-t_0)\}+r\times[\exp\{\lambda(t-t_0)\}-1]$$

若此画是真品，则 $t-t_0\approx 300$(年)。画中每克白铅所含铅210目前的分解数 $\lambda y(t)$及目前镭的分解数 r 均可用仪器测出，从而可求出 λy_0 的近似值，并利用(1)判断这样的分解数是否合理。若判断结果为不合理，则可以确定此画必是赝品，但反之不一定说明画是真品(因为估计仍是十分保守的，且只能证明画的“年龄”)。

(5)判断

Carnegie Mellon大学的科学家们利用上述模型对部分有疑问的油画作了鉴定，测得数据如下(见表4-6)。

表 4-6　几幅油画测定结果

油画名称	《在埃牟斯的门徒》	《濯足》	《看乐谱的女人》	《演奏曼陀林的女人》	《花边织工》	《笑女》
210 分解数（个/分）	8.5	12.6	10.3	8.2	1.5	5.2
镭 226 分解数（个/分）	0.8	0.26	0.3	0.17	1.4	6.0
计算 λy_0（个/分）	98050	157130	127340	102250	1274.8	−10181

判定结果：通过鉴定《在埃牟斯的门徒》得，$\lambda y_0 \approx 98050$[个/(克·分钟)]，它必定是一幅伪造品。类似可以判定《濯足》《看乐谱的女人》《演奏曼陀林的女人》也是赝品。而《花边织工》和《笑女》都不会是几十年内的伪制品，因为放射性物质已处于接近平衡的状态，这样的平衡不可能发生在 19 世纪和 20 世纪的任何作品中。

4.5　文物鉴定的科技方法

采用科技手段，利用现代科学仪器对文物进行鉴定的方法是多样的。比如：文物中的微量元素含量、元素的同位素比值等都是由原料矿物的形成条件决定的，就像人与生俱来的指纹一样，无法伪造，现代文物鉴定仪器可以通过测定这些数据，对文物进行指纹识别，从而判定它们的产地和年代。再如：文物艺术品的内部物质结构自诞生之日起就会随着时间推移不断自行调整，通过对其微观物质结构的测试分析，可以确定其结构变化的程度、推测其形成条件和年代。此外，采用高分辨率的显微成像系统，可以清晰地观测到器物的质地、微观形貌、加工痕迹和使用痕迹等特征，进而推测出器物的加工工艺、生产时代、器物功用等。

目前已应用于艺术品的科技鉴定手段主要有碳 14 断代、释光断代、成分检测、结构检测等。

4.5.1　碳 14 断代技术

利用放射原理，可以对许多文物的年代进行测定，例如对有机物(动植物遗体)的测定。考古学上目前流行的测定方法是放射性碳 14(^{14}C)测定法，这种方法具有较高的精确度，碳 14 是碳的一种具放射性的同位素，于 1940 年首次被发现。它是透过宇宙射线撞击空气中的氮原子而产生的，其半衰期约为 5730 年，衰变方式为 β 衰变，碳 14 原子转变为氮原子。有机物存活时，它们通过新陈代谢与外界进行物质交换，使体内的^{14}C 处于放射性平衡中。一旦有机物死亡，新陈代谢终止，放射性平衡即被破坏。因而，通过对比测定，可

以估计出它们生存的年代。

^{14}C 测年主要是利用C同位素^{14}C的稳定衰变来进行绝对时间的判断，测年时序范围在50000年以内，通过树木年轮的校正，采用最先进的质子加速技术，误差甚至可以控制在10年之内。在^{14}C技术出现以前，对考古遗址的绝对年代的判断几乎是不可能的事情，只能根据地层关系进行相对年代的划分，可以说^{14}C方法的建立具有极大的意义。此方法的建立者美国科学家W. F. Libby也因此获1960年诺贝尔化学奖。

著名的夏商周断代工程就是通过此法进行断代，确定了夏商周的帝王年表。又如，1950年在巴比伦发现一根刻有Hammurabi王朝字样的木炭，经测定，其^{14}C衰减数为4.09个/(克·分钟)，而新砍伐烧成的木炭中^{14}C衰减数为6.68个/(克·分钟)，由此可以推算出该王朝约存在于3900～4000年前。

遗憾的是，^{14}C测年对样品的要求是含C元素的有机质，比如陶瓷、金属器等无机材料，其C含量几乎可以忽略不计，就不能用^{14}C测年方法，所以采用^{14}C进行测年对于测试的艺术品的种类有一定的局限性。

4.5.2 释光断代技术

释光方法属于剂量学范畴，它与矿物晶体对辐射剂量的响应密切相关。其机理解释多借助固体能带理论，即认为是辐射电离导致矿物晶格中的电子产生跃迁并被陷阱俘获，其能量可以保存上百万年。当它受到人工或天然条件下的二次激发时，俘获电子得到新能量的补充再次激活，重新跃出陷阱并在发光中心与空穴复合回到基态，多余的能量以光子发射形式释放，称为释光。释光因其二次激发的方式不同，又可以进一步细分：以热激发的称之为热释光，以光激发的称之为光释光。目前已知自然界的多种矿物晶体都具有释光特性，在一定的辐照剂量范围内，矿物的释光量与其累积的吸收剂量呈正比，根据专门的实验方法可以推算出它们的等效剂量。

通过测定标本的天然放射性元素含量，可以计算出矿物自身及周边环境每年所能提供的辐射剂量，通常把这个辐射剂量称之为年剂量。因此，断代公式可以简单表达为：

标本距今年龄＝等效剂量/年剂量

即：标本中的矿物累积吸收剂量除以每年得到的辐射剂量，就可以推算出标本接受辐射的累计时间，即通常所说的样品释光断代年龄。显然，没有经历过核辐射作用的标本是不可能产生释光信号的。由于释光方法仅与长寿命放射性元素照射有关，而放射性衰变又是个常数，因此，释光断代年龄又被称为绝对年龄。意指该方法无需其他参照系，不受其他因素干扰，断代年龄可

以以精确数字形式表达。

释光断代方法是第四纪测年手段之一，主要原因有三：一是因为释光断代所用的标本多为地表松散的砂土颗粒，样品采集比较方便；二是因为释光断代的时间跨度可以从数百年到数十万年之间，甚至可以再上下各浮动一个数量级，基本涵盖了第四纪漫长的时域空间；三是因为该方法无需其他参照系，特别适合弥补黄土等“哑层”无年龄数据的尴尬。更详细内容参阅郑公望等人的《释光断代的可靠性分析》一文。

热释光强弱也被作为陶瓷类器物的计时标准，也适用于其他黏土制品。陶瓷在古代烧制时，经历几百到一千度高温，能将储存的能量以光的形式释放出来，这是放射性能量储存的标志。释放后又因继续放射性照射而积累，积累的能量多少与陶器烧成后经历的时间成正比。利用这种现象可以推算出陶器及其他黏土制品的烧制年代，一方面，陶瓷中放射性物质依然存在，另一方面，陶瓷埋藏处仍有放射性物质存在，并仍然会受宇宙射线的影响。于是，陶瓷中晶体又以均匀速率继续接受和贮藏辐射能，这些辐射能很“纯净”，是器物“诞生”后再增加的，可作为陶瓷器年龄的标志。

热释光断定陶瓷的真伪从方法上是可行的，但是有几点需要注意。一是这个方法需要取样，也就是说不能做到无损检测，鉴于陶瓷的不可恢复性，这样的有损检测应当尽量避免。二是释光法测试的时间跨度虽然很长，但是误差也很大，这样的误差通过经验系数校正，效果也不理想。三是可以通过人工辐射人为控制晶体贮藏的辐射能，进而影响测试结果。有研究表明，即使是太阳曝晒都会影响到陶瓷的热释光年代测定。

4.5.3 成分检测技术

成分检测是近年来发展起来的一种应用较为广泛的测试方法，在艺术品鉴定中的应用，主要是通过对比已知数据库中的真品数据，来判定未知样品的真伪。成分检测有多种手段和设备，例如湿化学法、INAA 法、ICP 法、XRF 法等。

(1)湿化学法(NPP 法)

狭义的湿化学法是指共沉淀；广义的湿化学法定义是有液相参加的、通过化学反应来制备材料的方法统称，如化学液相沉积(CBD)、电化学沉积(电镀)、溶胶凝胶等。

如甘油三酯检测(湿化学法)：从血清中抽提甘油三酯同时沉淀蛋白质，用氢氧化钾造化释出甘油，用过碘酸钠氧化成甲醛，然后用变色酸显色反应或用乙酰丙酮显色产生荧光色素，用荧光法或比色法测定。

(2)中子活化分析方法(INAA法)

中子活化分析方法(INAA法)利用中子照射某些元素,产生核反应,使这些元素转变为放射性核素的过程,称为活化。研究活化产生的核素的放射性特点,如半衰期、射线的种类和能量等,来确定试样中某些元素含量的方法。活化时,可以用快中子或慢中子照射,所以又分为快中子活化法和慢中子活化法。作为活化法的中子源有镭铍源、镍铍源、中子发生器及核反应堆等,用核反应堆产生的强大中子流及多道能谱仪,可以分析样品中某些极其微量的元素。小型中子管配合能谱仪,可以在井中测量某些元素的含量。仪器中子活化(INAA)具有分析灵敏度高、准确度高和取样量小等优点,被认为是最有效的古陶瓷微量元素分析方法,是研究古瓷产地特征的理想方法。

冯松林等人在论文《不同窑口古代青瓷产地特征的中子活化分析》中,用仪器中子活化有损分析了浙江寺龙口越窑、江西洪州窑和陕西耀州窑出土的古瓷碎片样品。通过对实验数据进行统计分析,结果表明3个窑口古瓷的元素组成具有各自的特征,它们之间具有可以区分的依据,寺龙口越窑与江西洪州窑的制胎原料关系较近,与北方的陕西耀州窑的关系较远,这些试验结果为古陶瓷地研究提供了重要参数。

(3)电感耦合等离子体法(Inductively Coupled Plasma, ICP)

电感耦合高频等离子体(Inductively Coupled Plasma,ICP)是原子光谱分析中一种广为采用的光源。当在感应线圈上施加高频电场时,由于某种原因(如电火花等)在等离子体工作气体中部分电离产生的带电粒子在高频交变电磁场的作用下做高速运动,碰撞气体原子,使之迅速、大量电离,形成雪崩式放电,电离的气体在垂直于磁场方向的截面上形成闭合环形的涡流,在感应线圈内形成相当于变压器的次级线圈并同相当于初级线圈的感应线圈耦合,这种高频感应电流产生的高温又将气体加热、电离,并在管口形成一个火炬状的稳定的等离子体焰炬。

等离子体法有ICP-AES和ICP-MS两种,前者是Inductively Coupled Plasma—Atom Emission Spectrometer的简写,是原子发射光谱仪;后者是Inductively Coupled Plasma—Mass Spectrometry的简写,是质谱仪。分析方法不一样,前者是原子外层的电子发出的特征光谱线经由光栅分光后来分析的;后者是电子在磁场中运行轨迹不一样、荷质比不一样来分析的。ICP法主要用于微量元素的分析,可分析的元素为大多数的金属和硅、磷、硫等少量的非金属,共72种。由于其检测灵敏度高,主要用于痕量分析,如动植物组织、食品、地质样品、水及环境样品中痕量元素的分析。广泛地应用于质量控制的元素分析、超微量元素的检测,尤其是在环保领域的水质监测。还可以对

常量元素进行检测，例如组分的测量中对主要成分的元素测定。

(4)X射线荧光光谱法(XRF法)

X射线荧光光谱仪(XRF)是一种较新型的可以对多元素进行快速同时测定的仪器。在X射线激发下，被测元素原子的内层电子发生能级跃迁而发出次级X射线(即X荧光)。波长和能量是从不同的角度来观察描述X射线所采用的两个物理量。波长色散型X射线荧光光谱仪(WD-XRF)是用晶体分光而后由探测器接收经过衍射的特征X射线信号。如果分光晶体和探测器作同步运动，不断地改变衍射角，便可获得样品内各种元素所产生的特征X射线的波长及各个波长X射线的强度。可以据此进行定性分析和定量分析。该种仪器产生于20世纪50年代，由于可以对复杂体系进行多组分同时测定，受到广泛关注，特别是地质部门先后配置这种仪器，分析速度显著提高。随着科学技术的进步，在20世纪60年代初发明了半导体探测器以后，对X荧光进行能谱分析成为可能。能谱色散型X射线荧光光谱仪(ED-XRF)，用X射线管产生原级X射线照射到样品上，所产生的特征X射线(荧光)直接进入Si(Li)探测器，便可以据此进行定性分析和定量分析。第一台ED-XRF在1969年问世，由于商品ED-XRF仪器及计算机软件的不断发展，功能逐步完善，应用领域拓宽，其特点、优越性被广泛关注，发展迅猛。因快速、精度高等特点在分析领域有着极为重要的地位，适合原材料、新材料、产品的研究、管理、分析，主要应用于水泥、钢铁、建材、石化、有色、硅酸盐、煤炭、高岭土、耐火材料、科研、环保等行业，可对多种样品进行从微量到高含量的分析。

在文物古玩鉴定领域，XRF最受青睐，民用化程度也最高。其测试速度快，无损，成本相对较低，并且有大量已发表的数据可查。但是这样一种测试方法也存在着一些无法忽视的问题。首先，成分检测不能直接断代，是一种间接证明测试样品真伪的手段，是一种必要非充分条件。成分检测和真品数据不符的样品必然不是真品，但是相符的样品也未必就是真品，这是成分检测最大的局限性。事实上，这种无损检测的方法只能作为一种参考依据，绝不能作为判断真伪的绝对标准。

4.5.4 结构测试技术

结构测试是根据陶瓷釉面或者胎体的结构变化，推导出陶瓷烧成的年代。目前已知的在陶瓷鉴定中应用比较成熟的结构检测方法是脱玻化检测。主要原理是：经高温熔融形成的釉子，是一种玻璃态均质体，其内部结构是无序的，在自然环境中，呈亚稳定状态。随着时间的推移，它的内部结构会不断自动地进行调整，由无序的亚稳定状态逐步向有序化的稳定状态转变，形成

微细晶体，这就是釉子的“脱玻璃化”现象。正是由于釉子存在这种自然“老化”现象，致使其内部结构随着时间的推移在不断发生变化，表现形式就是其透光性逐渐降低，对光线散射性不断增强。因此古陶瓷的釉面看上去要比新品柔和、温润，时代越久，这种反差就越大。有些品种的古陶瓷，这种“老化”现象表现得尤为突出，甚至造成釉面出现不同形态的微裂纹。脱玻化的过程几乎不受外界因素干扰，并且随着时间的推移，脱玻化的程度不断加深，较少受外界自然环境的影响。使用物理方法和化学方法检测，主要是会对釉子表面有一些损伤，对釉子的内部结构影响较少。其局限性表现在，脱玻化检测的对象只能是挂釉的陶瓷器，对于素胎无釉的器物是无法检测的，此外，如果是经过复烧的瓷器，脱玻化的检测结果一律为现代工艺品。

4.6 多管齐下才能更可靠

文物古玩存在极大的赢利空间，随之而来的也有一些不正常的现象，充斥着形形色色极端丑陋的故事。有人写道，文物古玩世界，因为这个浮躁的社会，人们内心尊崇敬仰的已经并非大师或大师的艺术成就，而是一群利欲熏心或沽名钓誉的家伙合伙打着艺术的旗号，举着大师的名牌，行招摇撞骗之实，甚至包括大师徐悲鸿的儿子，显而易见，谁又能不怀疑他是否同样在与拍卖行或其他专家们串通一气、坑蒙拐骗呢？这个社会似乎从来就不缺少钱多人傻或沽名钓誉的“收藏家”们，当然还有那些装模作样“鉴赏”艺术的伪艺术精英们。

现代书画家高鸿曾写过：“一般说来，书画收藏者和投资者对现代书画名家的名字，可谓是耳熟能详，但对他们作品的真伪甄别却并非都能做出正确判断。我曾困惑，为什么流通在书画市场上且离开我们时间并不长的书画家作品大多是赝品？而在我对一些拍品做真伪鉴定时，不时也会有感到非常棘手的无奈。通过长时间的调查和鉴定实践，似乎弄明白了其中的一点蹊跷，那就是：越是离我们近或我们越熟悉的现代书画家，其作品所显现的时代气息、个性风貌、笔墨习性对那些作伪者来说越容易领会、把握，更何况其中一些作伪者原本就是书画家的后人和嫡传弟子，加之书画家生前用过的印章俱在，其生前用过的材料现在也并不难找到，赝品的层出不穷以致几可乱真也就可想而知了。”

黑格尔在《法哲学原理》《小逻辑》中都提到：“凡合乎理性的东西都是现实的，凡现实的东西都是合乎理性的”，即所谓的“存在即合理”。在黑格尔那里，理性不仅仅是主观的理想性，而且是事物的本质，而事物是符合自己的本

质的，所以合乎理性的东西一定会成为现实；而只有合乎理性的东西才能“称”为现实，因而一切现实的东西就都是合理的。由于黑格尔认为，历史的发展就是绝对精神的发展和自我实现，而最为本体的绝对精神是好的，所以对于黑格尔而言，他就过滤了“现实”。他说：“在日常生活中，任何幻想、错误、罪恶以及一切坏东西，一切腐败幻灭的存在，尽管人们随便把它们叫做现实，但是，甚至在平常的感觉里，也会觉得一个偶然的存在不配享受现实的美名。因为所谓偶然的存在，只是一个没有什么价值的、可能存在的，亦即可有可无的东西。”黑格尔所言的现实，是经他重新定义过的，并非口语中平常意义的现实。

然而，“存在即合理”常被误用。面对大自然、面对社会，个人是非常渺小、无力的，个人没有能力改变这个世界，不管身边世界是多么丑陋，身边人们是多么低俗。丑恶的人和事存在必然有其存在的理由，“存在即合理”被误用也就不足为奇了。个人唯一能做的只有勇敢地面对现实。如果想投身于文物古玩的世界中，只能是将多样的科技测试手段应用到艺术品的鉴定中，与传统的眼学鉴定方法相结合，从而形成更加科学、客观、准确和完善的“多轨制”鉴定方法，这成为规避艺术品投资风险的重要手段，也在市场上被一些机构采用，服务于大众收藏者和投资者。在文物艺术品的投资中，坚持对文物艺术品的鉴定采用“多轨制”的鉴定方法，才能确保投资的文物艺术品得以保值和增值，屏蔽文物艺术品投资领域中的风险。

思考题

1. 经典武侠小说《射雕英雄传》是金庸于1957年到1959年创作的作品，在《香港商报》连载，边创作边连载，共40回。请在该小说中选择与情节无关的若干虚词（如：之、其、或、亦、……，呀、吗、的、着、是、在、可、便、就、但、……），统计它们在每一章出现的频数分布，分别以第1章、第5章、第10章为单位，研究这些词在这些单位中出现的频数是否相近？

2. 金庸武侠作品共15部，按出版顺序分别是《书剑恩仇录》《碧血剑》《射雕英雄传》《神雕侠侣》《雪山飞狐》《飞狐外传》《倚天屠龙记》《鸳鸯刀》《白马啸西风》《连城诀》《天龙八部》《侠客行》《笑傲江湖》《鹿鼎记》《越女剑》。请选择其中6部（要求早、中、晚期各2部），统计其中若干虚词出现频数的分布特征，对比前、中、后期，分析作者创作用词方式会否受时间的影响而发生变化，如何变化？

3. 马克·吐温作品丰富，被誉为“第一位真正的美国作家，我们都是继承他而来”。在美国，马克·吐温似乎有两个，一个是读者眼中的马克·吐温，

一个是批评家眼中的马克·吐温。在读者看来,马克·吐温幽默、滑稽、诙谐,夸张得要命,可又蛮真实的,有时还带点刺儿;他的美国味道很足,男女老少都喜欢读。在批评家看来,事情就复杂了。请统计分析马克·吐温《百万英镑》的用词情况。

第5章

社会历史演变规律——人口模型

中国历史，分合交替，分久必合，合久必分。夏一统，商周继之，春秋战国乱之；秦一统，两汉继之，三国魏晋南北朝乱之；隋一统，大唐继之，五代十国宋辽金乱之；元一统，明清继之，民国乱之。从表面上看，国家兴亡盛衰主要取决于帝王的才能和政策取向。皇帝圣明勤政、任人唯贤，大小官员基本能奉公守法，都不敢以权谋私、胡作非为，人民能安居乐业，民风敦厚，国家呈现民富国强的欣欣向荣的局面。反之，皇帝昏庸无能、任人唯亲，大小官员都以权谋私、违法乱纪，甚至买卖官职、胡作非为，把国家弄得两极分化、民不聊生，最后不是皇亲国戚争权夺利就是权党搞宫廷政变，或者官逼民反，进行新一轮改朝换代的战乱，直至出现一位能因势利导的开国帝王。本章目的是通过描述地区人口演变规律，试图揭示其对社会历史演变的客观影响。

先以唐朝为例。公元755年(天宝十四年)的十月、十一月之间，杜甫自京赴奉先县写下千古绝句“朱门酒肉臭，路有冻死骨”。唐代经历唐太宗的“贞观之治”、唐高宗的“永徽之治”、武则天的“武周之治”及唐玄宗的“开元之治”后，国势大增，文治武功在唐玄宗开元年间达至鼎盛状态，属史无前例的盛世。如此盛世居然“路有冻死骨”，说明贫富分化已非常严重。据对历史上人口的研究，当时唐朝人口至少已达7500万人，约是今天13.7亿的1/18。在距今1250年前的唐朝，农业生产是非常低下的，如此多人口已属严重超负荷。另一方面，唐朝经过100多年的平稳发展，社会财富越来越集中在少数人手中，老百姓饿死、冻死就不足为奇了。

5.1 马尔萨斯人口理论

5.1.1 马尔萨斯生平

托马斯·罗伯特·马尔萨斯牧师(Thomas Robert Malthus，1766—1834)是英国人口学家和政治经济学家，他的学术思想悲观，但影响深远。

马尔萨斯出生于一个富有的家庭，年幼时在家接受教育，直到1784年被剑桥大学耶稣学院录取。他在那里学习了许多课程，并且在辩论、拉丁文和

希腊文课程中获奖，主修科目是数学。1791 年他获得硕士学位，并且在两年后当选为耶稣学院院士。1797 年他被按立为圣公会的乡村牧师。1798 年，马尔萨斯针对 18 世纪末英国工业革命造成的大批工人失业、贫困问题突出等社会问题，以匿名的方式发表了自己的《人口原理》，其理论与 19 世纪盛行的社会富足促成的经济进步的乐观主义思想相抵触，因此遭到了猛烈批判。

1805 年，他成为英国第一位政治经济学教授，执教于东印度公司学院。他的学生亲切地称呼他为“人口”马尔萨斯。从那时直至去世，马尔萨斯一直都在黑利伯瑞的东印度公司学院担任政治经济学和近代史教授。

5.1.2 马尔萨斯人口理论

马尔萨斯的《人口原理》从两个抽象前提出发：第一，食物为人类生存所必需；第二，两性间的情欲是必然的，但几乎保持现状。认为在这两者中，人口增殖力比土地生产人类生活资料力更为巨大。在无所妨碍时，人口以几何级数率增加，即以 1、2、4、8、16、32、64、128、256、512 的增加率增加；生活资料将以 1、2、3、4、5、6、7、8、9、10 的算术级数增加率增加。因此生活资料的增加赶不上人口的增长是自然的、永恒的规律，只有通过饥饿、繁重的劳动、限制结婚以及战争等手段来消灭社会“下层”，才能削弱这个规律的作用。把资本主义制度造成的一切问题和灾难归结为人口过剩的结果。

《人口理论》第二章写道：“与欧洲任何国家相比，美国的生活资料一直较为充裕，风俗较为淳朴，所以也较少地抑制早婚。正因为如此，我们发现，该国的人口每 25 年就翻一番。”马尔萨斯的人口原理的主要论点之一，他以“土地收益递减规律”作为自己人口理论的经济学和自然史的基础，认为由于“土地收益递减规律”的作用，使食物赶不上以几何级数速度增加的人口的需要。

5.1.3 影响

在东印度公司学院，马尔萨斯发展出一套需求供应失衡理论，他称之为过剩。在当时这被看做荒唐的理论，却是后来有关大萧条的一系列经济理论的先驱，他的崇拜者、经济学家约翰·梅纳德·凯恩斯将这个思想引入了著作。以往认为高出生率有利于经济，因为会提供更多的劳动力。然而，马尔萨斯却从一个新的视角看待出生率，并且说服了大多数经济学家：即使高出生率可以增加毛生产量，它更趋于降低人均生产量。马尔萨斯具有广泛的影响力，他的崇拜者包括知名的经济学家大卫·李嘉图等人。

对马尔萨斯理论的关注也帮助了英国全国人口普查的实施。1801 年，政府官员约翰·李克曼主导了第一次现代人口普查。马尔萨斯人口理论的门

徒，还包括知名的神创论者、自然神学家威廉·佩里大主教，他于1802年发表了《自然神论》，他认为马尔萨斯的人口学原理证明了神的存在。

马尔萨斯学者对经济学也有着重要的影响。英国著名的经济学家大卫·李嘉图说："劳动的自然价格就是必须使劳动者能够共同生存，即使人类不增不减永世长存的价格。"这个学说一般被称为"工资钢铁定律"，为马克思接受，成为他剩余价值学说的一个主要成分。

马尔萨斯的观点还影响着生物学的研究。查理·达尔文在他的《物种起源》一书中说，他的理论是马尔萨斯理论在没有人类智力干预的一个领域里的应用。华莱士称马尔萨斯的著作是"我所阅读过的最重要的书"，并把他和达尔文通过学习马尔萨斯理论、各自独立地发展出进化论的事，称作"最有趣的巧合"。

马尔萨斯对于人口问题的思考是现代进化理论的基础，他强化了对"有限增长"条件下"生存挣扎"的观察。由于马尔萨斯理论，达尔文认识到了生存竞争不仅发生在物种之间，而且也在同一物种内部进行。联合国人口基金会等国际组织关于地球能容纳多少人的辩论即起源于马尔萨斯。

5.1.4 马尔萨斯模型

基本假设：在人口自然增长过程中，净相对增长（出生率与死亡率之差）是常数，即单位时间内人口的增长量与人口成正比，比例系数设为 r。如果以年为单位，则 r 代表人口数年平均增长率。在此假设下，推导并求解人口随时间变化的数学模型。

设时刻 t 的人口为 $N(t)$，把 $N(t)$ 当做连续、可微函数处理（因人口总数很大，可近似地这样处理，此乃离散变量连续化处理），根据马尔萨斯的假设，在 t 到 $t+\Delta t$ 时间段内，人口的增长量为

$$N(t+\Delta t)-N(t)=rN(t)\Delta t$$

并设 $t=t_0$ 时刻的人口为 N_0，于是

$$\begin{cases}\dfrac{\mathrm{d}N}{\mathrm{d}t}=rN\\N(t_0)=N_0\end{cases}$$

这就是马尔萨斯人口模型，用分离变量法易求出其解为

$$N(t)=N_0\mathrm{e}^{r(t-t_0)}$$

此式表明人口以指数规律随时间无限增长。另外，人口翻一番即增加1倍的时间间隔 $t-t_0=\ln 2/r$。

按照马尔萨斯的理论，每隔25年人口数量翻一番，则 $\ln 2/r=25$，即 $r=0.0277=2.77\%$。换句话说，如果以年均2.77%的速度增长，大约25年后人

口要增加 1 倍。

例 5.1 表 5-1 是美国 1790—1900 年每隔 10 年的人口数据。

表 5-1 1790—1900 年美国人口数据(单位:百万)

年份	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
人口	3.90	5.27	7.18	9.57	12.83	16.99	23.07	31.28	38.42	49.92	62.69	76.09
预测	4.17	5.49	7.22	9.50	12.50	16.44	21.64	28.47	37.46	49.29	64.85	85.33
误差	−0.27	−0.22	−0.04	0.07	0.34	0.54	1.43	2.81	0.96	0.64	−2.16	−9.24
每 10 年年均增长		0.030	0.031	0.029	0.029	0.028	0.031	0.030	0.021	0.026	0.023	0.019

记 t 年份的美国人口为 $N(t)$,假设美国人口按照马尔萨斯人口模型:$N(t)=ae^{bt}$,其中 a、b 为待定的未知参数。最理想的结果自然是所有实际数据点都恰好落在曲线上,但历史经验告诉我们,理论模型都只是实际数据的某种近似,我们的任务是找出误差尽可能少的理论模型。通常人们把实际数据称作观测值 $N(t)$,理论模型对应的值称为预测值 ae^{bt},观测值与预测值之间的差称为残差或误差 $N(t)-ae^{bt}$。为求得 a、b,考虑让总的误差达到最小。即

$$Q(a,b)=[N(t_1)-ae^{bt_1}]^2+[N(t_2)-ae^{bt_2}]^2+\cdots+[N(t_n)-ae^{bt_n}]^2$$

或

$$|N(t_1)-ae^{bt_1}|+|N(t_2)-ae^{bt_2}|+\cdots+|N(t_n)-ae^{bt_n}|$$

达到最小,一般是前一个式子即残差的平方和 $Q(a,b)$ 取到最小,这一方法叫做最小二乘法。当我们取遍所有 a 和 b,必然能找到一对 a、b 的值使 $Q(a,b)$ 达到最大,通过运算得 $a=1.93\times10^{-21}$ 和 $b=0.027$。也就是说,此时对应的估计(预测)值的总(平方)误差最小。理论预测值和误差见表 5-1 的第 3、4 行,预测拟合曲线见图 5-1。

另外,人口数每隔 10 年的年平均增长率计算如下:$r=\ln(N_2-N_1)/10$,结果见表 5-1 最后一行。1860 年前的美国人口年均增长率保持在 3%左右的较高水平,之后年均增长率明显下降,分别为 2.1%、2.6%、2.3%、1.9%。1861—1865 年美国南北战争是主要原因,战争导致人口增长率下降。南北战争后至第一次世界大战不到 50 年的时间内,美国从一个农村化的国家变成了城市化的国家,机器代替了手工,产品大量增加。工业化带来了人口出生率的下降,最终使得自然增长率下降。

例 5.2 世界人口:据估计,1961 年地球上的人口总数为 3.06×10^9,而在随后 7 年中,人口总数以每年 2%的速度增长,这样 $t_0=1961$,$N_0=3.06\times$

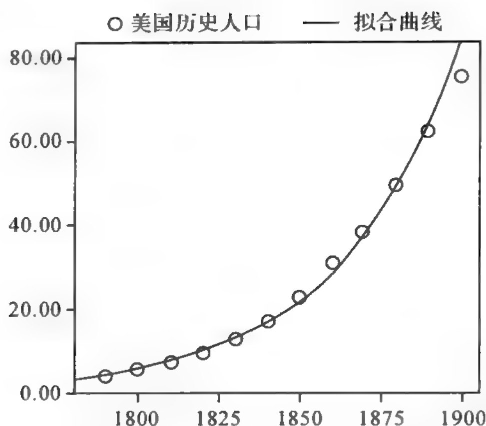


图 5-1　美国历史人口与拟合曲线

10^9，$r=0.02$，于是

$$N(t)=3.06\times10^9 e^{0.02(t-1961)}$$

这个公式非常准确地反映了在 1700—1961 年间世界人口总数。因为，这期间地球上的人口大约每 35 年翻一番，而上式断定 34.6 年增加一倍。

但是，后来人们以美国人口为例，用马尔萨斯模型计算结果与人口资料比较，却发现有很大的差异，尤其是在用此模型预测较遥远的未来地球人口总数时，发现更令人不可思议的问题，如按此模型计算，到 2670 年，地球上将有 36000 亿人口。如果地球表面全是陆地（事实上，地球表面还有 80%被水覆盖），我们也只得互相踩着肩膀站成两层了，这是非常荒谬的。因此，这一模型应该修改。

5.2　逻辑斯蒂克模型

5.2.1　逻辑斯蒂克模型

马尔萨斯模型为什么不能预测未来的人口呢？这主要是地球上的各种资源只能供一定数量的人生活，随着人口的增加，自然资源环境条件等因素对人口增长的限制作用越来越显著。如果当人口较少时，人口的自然增长率可以看做常数的话，那么当人口增加到一定数量以后，这个增长率就要随人口的增加而减小。因此，应对马尔萨斯模型中关于净增长率为常数的假设进行修改。

1838 年，荷兰生物数学家韦尔侯斯特（Verhulst）引入常数 N_m，用来表示自然环境条件所能容许的最大人口数（一般说来，一个国家工业化程度越高，

它的生活空间就越大，食物就越多，从而 N_m 就越大），并假设将增长率等于 $r[1-N(t)/N_m]$，即净增长率随着 $N(t)$ 的增加而减小，当 $N(t)\to N_m$ 时，净增长率趋于零，按此假定建立人口预测模型。

由韦尔侯斯特假定，马尔萨斯模型应改为

$$\begin{cases}\dfrac{\mathrm{d}N}{\mathrm{d}t}=r\left(1-\dfrac{N}{N_m}\right)N,\\ N(t_0)=N_0\end{cases}$$

上式就是逻辑模型，该方程可分离变量，其解为

$$N(t)=\frac{N_m}{1+\left(\dfrac{N_m}{N_0}-1\right)\mathrm{e}^{-r(t-t_0)}}$$

下面，我们对模型作一简要分析：

(1)当 $t\to\infty$，$N(t)\to N_m$，即无论人口的初值如何，人口总数趋向于极限值 N_m。

(2)当 $0<N<N_m$ 时，$\mathrm{d}N/\mathrm{d}t=r(1-N/N_m)N>0$，这说明 $N(t)$ 是时间 t 的单调递增函数。

(3)由于 $\mathrm{d}^2N/\mathrm{d}t^2=r(1-2N/N_m)$，所以当 $N<N_m/2$ 时，$\mathrm{d}^2N/\mathrm{d}t^2>0$，$\mathrm{d}N/\mathrm{d}t$ 单调递增；当 $N>N_m/2$ 时，$\mathrm{d}^2N/\mathrm{d}t^2<0$，$\mathrm{d}N/\mathrm{d}t$ 单调递减。即人口增长率 $\mathrm{d}N/\mathrm{d}t$ 由增变减，在 $N_m/2$ 处最大，也就是说在人口总数达到极限值一半以前是加速生长期；过这一点后，生长的速率逐渐变小，并且迟早会达到零，这是减速生长期。

(4)用该模型检验美国从 1790 年到 1950 年的人口，发现模型计算的结果与实际人口在 1930 年以前都非常吻合，自从 1930 年以后，误差愈来愈大，一个明显的原因是在 20 世纪 60 年代美国的实际人口数已经突破了 20 世纪初所设的极限人口。由此可见该模型的缺点之一是 N_m 不易确定，事实上，随着一个国家经济的腾飞，它所拥有的食物就越丰富，N_m 的值也就越大。

(5)用逻辑模型来预测世界未来人口总数。某生物学家估计，$r=0.029$，又当人口总数为 3.06×10^9 时，人口每年以 2% 的速率增长，由逻辑模型得

$$\frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}t}=r\left(1-\frac{N}{N_m}\right),$$

即

$$0.02=0.029\left(1-\frac{3.06\times10^9}{N_m}\right),$$

从而得

$$N_m=9.86\times10^9,$$

即世界人口总数极限值近 100 亿。

值得说明的是，人也是一种生物，因此上面关于人口模型的讨论，原则上也可以用于在自然环境下单一物种生存着的其他生物，如森林中的树木、池

塘中的鱼等,逻辑模型有着广泛的应用。

5.2.2 再论美国历史人口

例 5.3 表 5-2 是美国 1790—2000 年每隔 10 年的人口数据,更详细的人口数据可以到美国人口普查局网站(http://www. census. gov/popest/national/)查阅。求其经验回归方程。

表 5-2 1790—2000 年美国人口数据(单位:百万)

年份	人口	年份	人口	年份	人口	年份	人口	年份	人口
1790	3.895	1840	16.985	1890	62.692	1940	132.122	1990	249.464
1800	5.267	1850	23.069	1900	76.094	1950	152.271	2000	282.124
1810	7.182	1860	31.278	1910	92.407	1960	180.671		
1820	9.566	1870	38.416	1920	106.461	1970	205.052		
1830	12.834	1880	49.924	1930	123.076	1980	227.224		

对表 5-2 中数据利用逻辑斯蒂克模型进行拟合,分别假设美国人口上限为 $N_m=350$(百万)和 $N_m=400$(百万),拟合效果差异不大,见图 5-2;模型公式分别如下:

$$N(t)=350/[1+98.62e^{0.02634(t-1790)}]$$

和

$$N(t)=400/[1+102.70e^{-0.02532(t-1790)}]$$

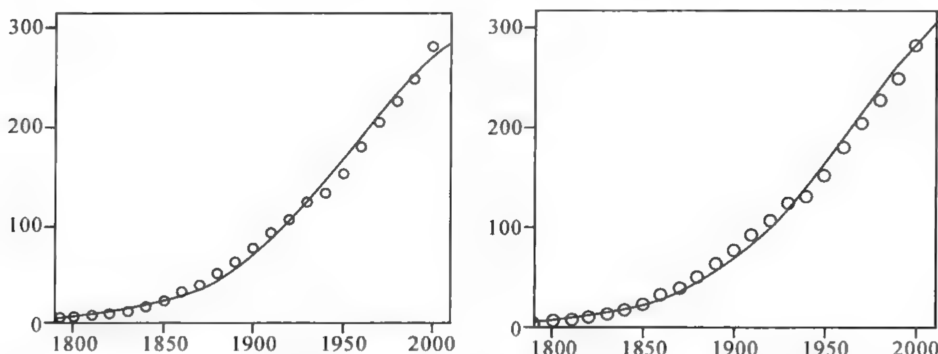


图 5-2 美国历史人口用逻辑斯蒂克模型拟合(左 $N_m=350$,右 $N_m=400$)

计算每 10 年的年平均增长率,见表 5-3 和图 5-3。我们发现可将美国人口增长率分成 3 个阶段:1790—1860 年、1860—1910 年、1910—2010 年。

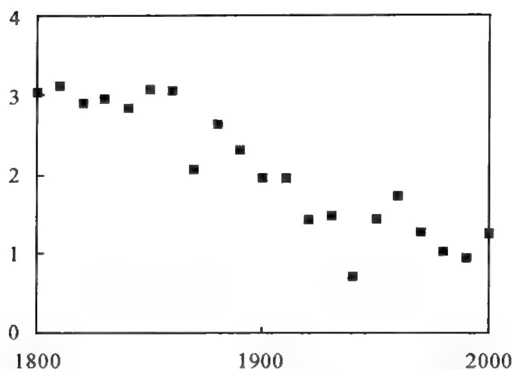


图 5-3 美国历史人口年均增长率(%)

表 5-3 1790—2000 年美国人口年平均增长率(单位:%)

年份	年增长率	年份	年增长率	年份	年增长率	年份	年增长率	年份	年增长率
1790		1840	2.802	1890	2.277	1940	0.709	1990	0.934
1800	3.018	1850	3.062	1900	1.937	1950	1.419	2000	1.230
1810	3.101	1860	3.044	1910	1.942	1960	1.710		
1820	2.866	1870	2.056	1920	1.416	1970	1.266		
1830	2.939	1880	2.620	1930	1.450	1980	1.027		

(1)1790—1860 年,美国建国的初创时期,以农业生产为主,人口年平均增长率保持在 3%左右。

(2)1860—1910 年,改革和工业化时期,以工业生产为主,人口年平均增长率从前期的 3%左右逐步下降到约 2%。早在 19 世纪初期,美国就开始了工业化,但传统农业、传统文化仍占主导地位。1861—1865 年美国南北战争是关键的转折点,内战以代表新生事物的北方战胜代表落后势力的农奴制的南方而结束,美国开始进入新的发展时期,工业化逐渐步入成熟阶段。在从美国内战至第一次世界大战前的大约 50 年时间内,美国从一个农村化的共和国变成了城市化的国家,机器代替了手工,产品大量增加。全国性的铁道网形成,增进了货品流通;应大众的需要,许多新发明出现并很快被投放市场;银行业提供贷款,促成工商业经营的扩大。其中 1890 到 1917 年的近 30 年间被称为所谓的“进步时期”。1861—1870 年,因为国内战争,人口年平均自然增长率几乎下降了整整 1 个百分点,此后随着工业化的推进,人口增长率逐步下降。

(3)1910—2000 年,美国在世界舞台上开始扮演重要角色。人口年平均

增长率基本保持在1.5%以下。1914年,第一次世界大战爆发。1917年,美国终于被卷入大战漩涡中。1910—1920年,因为第一次世界大战,美国人口年增长率下降到1.5%以下。除了50年代曾达到1.7%外,其他时间的年增长率都低于1.5%。

5.2.3 市场价格模型

对于纯粹的市场经济来说,商品市场价格取决于市场供需之间的关系,市场价格能促使商品的供给与需求相等,这样的价格称为(静态)均衡价格。也就是说,如果不考虑商品价格形成的动态过程,那么商品的市场价格应能保证市场的供需平衡,但是,实际的市场价格不会恰好等于均衡价格,而且价格也不会是静态的,应是随时间不断变化的动态过程。

假设在某一时刻 t 商品的价格为 $p(t)$,它与该商品的均衡价格间有差别,此时存在供需差,此供需差促使价格变动,对新的价格,又有新的供需差,如此不断调节,就构成市场价格形成的动态过程。假设价格 $p(t)$ 的变化率 $\frac{\mathrm{d}p}{\mathrm{d}t}$ 与需求和供给之差成正比,并记 $f(p,r)$ 为需求函数,$g(p)$ 为供给函数(r 为参数),于是

$$\begin{cases}\dfrac{\mathrm{d}p}{\mathrm{d}t}=\alpha[f(p,r)-g(p)],\\ p(0)=p_0\end{cases}$$

其中 p_0 为商品在 $t=0$ 时刻的价格,α 为正常数。

若设 $f(t,r)=-ap+b, g(p)=cp+d$,则上式变为

$$\begin{cases}\dfrac{\mathrm{d}p}{\mathrm{d}t}=-\alpha(a+c)p+\alpha(b-d),\\ p(0)=p_0\end{cases}$$

其中 a,b,c,d 均为正常数,其解为

$$p(t)=\left(p_0-\frac{b-d}{a+c}\right)\mathrm{e}^{-\alpha(a+c)t}+\frac{b-d}{a+c}$$

下面对所得结果进行讨论:

(1)设 $\bar{p}$ 为静态均衡价格,则其应满足

$$f(\bar{p},r)-g(\bar{p})=0,$$

即 $-a\bar{p}+b=c\bar{p}+d$,于是得 $\bar{p}=\frac{b-d}{a+c}$,从而价格函数 $p(t)$ 可写为

$$p(t)=(p_0-\bar{p})\mathrm{e}^{-\alpha(a+c)t}+\bar{p}$$

令 $t\to+\infty$,取极限得

$$\lim_{t\to+\infty}p(t)=\bar{p}$$

这说明，市场价格逐步趋于均衡价格。又若初始价格 $p_0=\bar{p}$，则动态价格就维持在均衡价格 $\bar{p}$ 上，整个动态过程就化为静态过程。

(2)由于

$$\frac{\mathrm{d}p}{\mathrm{d}t}=(\bar{p}-p_0)\alpha(a+c)\mathrm{e}^{-\alpha(a+c)t}$$

所以，当 $p_0>\bar{p}$ 时，$\frac{\mathrm{d}p}{\mathrm{d}t}<0$，$p(t)$ 单调下降向 $\bar{p}$ 靠拢；当 $p_0<\bar{p}$ 时，$\frac{\mathrm{d}p}{\mathrm{d}t}>0$，$p(t)$ 单调增加向 $\bar{p}$ 靠拢。这说明：初始价格高于均衡价格时，动态价格就要逐步降低，且逐步靠近均衡价格；否则，动态价格就要逐步升高。因此，模型在一定程度上反映了价格影响需求与供给，而需求与供给反过来又影响价格的动态过程，并指出了动态价格逐步向均衡价格靠拢的变化趋势。

5.3 中国历史人口演变

中国人口史，是研究中国历史上的人口组成、变迁、分布等的分支学科。中国是目前世界上人口最多的国家，在历史上中国的人口数量在相当长的时间内占有世界很高的比例。虽然中国历史上从周宣王时期就有人口调查，在战国前期就有详细的人口统计，可惜文献资料没有流传下来，或由于各种原因的影响，多不精确。但自班固编纂《汉书》开始，之后历代 16 本正史中的《地理篇》大多有人口记载，均为研究人口变迁的珍贵资料。

5.3.1 人口资料的获得和人口的估算

人口资料多见于正史中记录的纳赋人口，但是为了逃避赋税等，为数不少的人口并不见于正史的记录，而且有很多和尚、道士、奴婢及部曲等也并不归于此类，造成统计上的偏差。在战乱当中，国家控制力下降，人口逃逸，所以往往人口数据需要加入估算才更接近于真实。

(1)春秋战国时代

春秋末年到战国初年的这一时期(前 481 年—前 453 年)是古代中国传统农业的迅速发展和奠定后来传统农耕文明基础的时代，期间各诸侯国的人口以较快速度稳定增长。《国语》记载，周宣王时中国就已进行过人口调查。真正详细的人口调查起源于战国初期，但当时的资料并没有流传下来。估计在战国后期的前 230 年，根据当时各国兵力及资料推算，中原各国人口总计 3000 万。此后，一直到秦末民变之前都保持持续增长。

(2)秦、汉

秦始皇二十六年(前 221 年)最终完成统一六国大业，之后持续不断的领

土扩张，到秦始皇三十七年(前 210 年)，全国总人口约 3 千万。估计此年的世界人口有 2 亿，秦代人口占世界总人口的 15%。秦末民变和楚汉战争造成期间全国的社会经济生产受到严重破坏，大量人口死亡。

公元前 202 年刘邦称帝，建立汉朝，人口约 1650 万，占世界人口的 8.66%。西汉建立后由于奉行黄老政治、与民生息的政治措施，到汉武帝元光六年时(前 129 年)人口达到 3600 万。之后连年的战争和徭役赋税较之前加重，使得自汉武帝在位中期开始，人口数出现了多年的停滞和下降。之后昭宣中兴的休养生息，人口开始恢复增长。到新莽天凤五年(公元 18 年)全国有 5600 万人口。但随后爆发绿林赤眉起义，西汉灭亡。汉光武帝建武元年(25 年)时估计全国有人口 2800 万；汉和帝元兴元年(105 年)约 5325 万；此后增长停滞，汉灵帝光和七年(184 年)约 5500 万人。

(3)魏、晋、南北朝

自东汉末黄巾之乱(184 年)开始，进入长期战乱时期。根据传世文献资料记载，到蜀汉灭亡时(263 年)人口合计在 820 万左右。可见东汉末年军阀混战之惨烈，这也从一个侧面为经典著作《三国演义》提供了素材。自此人口增长态势一蹶不振，到隋文帝在位时方复苏。

晋武帝太康元年(280 年)灭吴统一全国后，282 年户口调查为全国 377 万户、约 2380 万人。西晋的统一局面只维持了 11 年，就发生了“八王之乱”，紧接着发生了历史上著名的导致华北地区人口在 8 年之内消失 80%的“五胡乱华”。

南北朝国家分裂、社会动荡，人口数量在低位徘徊。

(4)隋、唐、五代

隋文帝开皇元年(581 年)建立隋朝，估计实际人口为 1800 万。隋炀帝大业五年(609 年)约 46,019,956 人。

唐太宗贞观二年(628 年)约 1900 万人，唐玄宗开元十四年(726 年)4141 万人，唐玄宗天宝十三年(754 年)约 7500 万人。安史之乱结束后，人口分布格局发生重大变化。

五代十国，四分五裂，人口不振。估计 960 年的五代十国人口总数为 6000 万。

(5)宋、元

宋朝时水稻在南方广泛种植，人口数量开始跃上新台阶。宋徽宗大观四年(1110 年)估计实际人口大约 12500 万人，其中华南地区人口占全国 70%、华北地区占 30%。辽国人口最多时的辽天祚帝乾统十年(1110 年)约 900 万人。西夏人口最多时的西夏崇宗贞观十年(1110 年)估计约有 300 万人。

绍兴和议后(1141 年)，从靖康之难开始因频繁战乱遭受破坏的社会经济

开始得到恢复，人口总量得到一定程度的恢复增长。到金章宗泰和七年(1207年)时，宋、金、西夏三国人口总数估计达到13600万人，约占当时世界人口的33.66%。

1205年，蒙古帝国首次进攻西夏，到1227年灭西夏，直到1279年灭南宋统一天下，人口跌至谷底。元代人口最高曾近9000万，由于战乱，到1368年元朝灭亡时，人口约6000余万。

(6)明、清

明太祖洪武二十六年(1393年)有6500万人，明宪宗成化十六年(1480年)估计实际人口有7185万。明世宗嘉靖末年美洲高产作物传入后开始在当时人口最为稠密的江浙和岭南地区普及和推广，到明思宗崇祯十三年(1640年)估计达到17500万人。明思宗崇祯十三年(1640年)到清世祖顺治七年(1650年)，爆发一种迄今为止未明的瘟疫，造成京杭大运河沿线和南直隶这一全国人口最稠密和经济发展重心地区的很多地方人口的50%消失，有的地方甚至高达80%。

清圣祖康熙二十四年(1685年)有10171万人，康熙六十一年(1722年)，上升到15000万人，清文宗咸丰元年(1851年)更是达到43216万人。随后爆发的太平天国及外国列强的侵略，人口增长停滞不前。

(7)中华民国

民国二十八年(1939年)，人口为517568000，约占世界人口的1/5。中国人口地理学创始人胡焕庸1935年划定黑河瑷珲—腾冲人口线(又称胡焕庸线)，在领土面积上，该线东南部分占全国的43%，西北部分占全国的57%；而在人口上，东南部分人口占全国的90%以上，西北部分人口却不足全国的10%。四川、江苏、山东为全国人口最多的三省。江苏、山东、浙江为人口密度最高的三省。

(8)新中国

新中国成立后共进行六次人口普查，分别为1953年59435万人，1964年69458万人，1982年10亿人，1989年11.3368亿人，2000年126583亿人，2010年137053亿人。

5.3.2 中国历史人口增长规律

根据上节资料，综合其他文献，表5-4列出相对认为合理的若干年份历史人口估计数，并计算出相应区间段的年均增长率。我们可总结出中国历史人口的变化规律有如下几点。

(1)在新朝代建立后进入数十年稳定发展的初期，因人少地多，朝廷通常

会采取休养生息政策，人口会快速增加，年均增长率可持续保持在 10‰以上，个别时期可高达 20‰甚至 30‰以上。秦、隋两朝都没有实行休养生息政策，结果变成短命王朝。

(2)随后社会进入鼎盛时期，但人口却停滞不前，人口与食物矛盾开始显现。杜甫写于安史之乱前夕的绝唱“朱门酒肉臭，路有冻死骨”，就是最佳写照，表面上看是辉煌的开元盛世，同时京城长安还有穷人饿死、冻死。

(3)一些貌似不大的事件却引发社会大动荡，最终旧皇朝分崩离析，人口大量死亡，人口数降至谷底，一直到新王朝的建立，开始新一轮的循环。如果在此期间有其他外来干扰，或不按常规行事，就可能缩短或延长波段时间。

表 5-4　中国特殊年份历史人口和年均增长率(‰)

年份	万人	增长率	年号或事件	年份	万人	增长率	年号或事件
前 221	3000			1200	13600	0.358	宋宁宗庆元六年
前 202	1650	−31.465	汉太祖元年	1293	6304	−8.268	元世祖至元三十年
前 129	3600	10.687	汉武帝元光六年	1330	8900	9.321	元文宗至顺元年
2	5767	3.597	汉平帝元始二年	1368	6000	−10.376	元朝灭亡
25	2800	−31.414	光武帝建武元年	1393	6500	3.202	明太祖洪武二十六年
105	5325	8.035	汉和帝元兴元年	1487	7500	1.522	明宪宗成化二十四年
184	5500	0.409	黄巾之乱开始	1640	17500	5.538	明思宗崇祯十三年
263	820	−24.091	蜀汉灭亡	1685	10171	−12.059	清圣祖康熙二十四年
282	2380	56.082	晋武帝太康三年	1722	15000	10.500	清圣祖康熙六十一年
421	2000	−1.251	南朝宋武帝永初二年	1850	43000	8.228	清宣宗道光三十年
440	2370	8.934	南朝宋文帝元嘉十七年	1887	37763	−3.510	清德宗光绪十三年
589	4000	3.513	隋文帝开皇九年	1901	45000	12.524	清德宗光绪二十七年
609	4601	6.999	隋炀帝大业五年	1939	51756	3.681	
628	1900	−46.548	唐太宗贞观二年	1953	59435	9.882	
754	7500	10.897	唐玄宗天宝十三年	1964	69458	14.167	
859	8000	0.615	唐宣宗大中十三年	1982	100818	20.700	
907	7500	−1.345	唐哀帝天复三年	1989	113368	16.760	
960	6000	−4.210	五代十国	2000	126583	10.024	

续表

年份	万人	增长率	年号或事件	年份	万人	增长率	年号或事件
1041	10000	6.306	宋仁宗庆历元年	2005	130628	6.291	
1125	13240	3.341	宋徽宗宣和七年	2010	137053	9.603	

5.4 世界历史人口演变

由于中国地理环境独特，东面、东南面临大海，西面、西南面有大山，北面是草原、沙漠。在漫长的冷兵器时代，这一切就是天然屏障，因此中国是合了分、分了又合，演绎着完整的发展史。在亚欧大陆和非洲大陆，现今没有一个国家的发展史像中国这么完整。由于世界历史的复杂性和史料明显缺乏，世界历史人口史的背景就不在本书探讨，只列出一些历史人口数据，进行简单解读。

5.4.1 人口资料的获得和人口的估算[①]

估计公元前1000年时的世界人口有5000万。估计公元前500年时的世界人口有1亿，波斯阿契美尼德帝国此年有2500万人口。估计公元前400年时的世界人口有1.62亿。估计公元前200年时的世界人口有1.905亿。估计公元前265年的印度孔雀王朝有5000万人口。

估计公元元年(1年)的世界人口为2.7227亿，汉代人口占世界的比例是21.18%。罗马帝国在前25年有270万平方公里土地、5680万人口；117年的罗马帝国有590万平方公里土地、8800万人口；395年的东罗马帝国有250万平方公里土地、3400万人口。

估计公元200年的世界人口为2.23亿，此年的东汉人口占世界比例的15.7%。

世界人口在950年—1200年间从2.5亿增加到4.04亿，拜占庭帝国在1025年有1200万人。

世界人口从1830年的10亿、1930年20亿、1960年30亿、1974年40亿、1988年50亿、2000年60亿，到2009年5月30日已经达到68亿。

---

① 1700年及之前世界及各大洲人口资料，来源于(意)马西姆·利维巴茨著:《繁衍:世界人口简史》(第3版)，郭峰、庄瑾译，北京大学出版社2005年版；1750年及之后资料，来源于互联网Wikipedia网站。

表 5-5　世界及各大洲历史人口(百万人)

年份	世界	亚洲	欧洲	非洲	拉丁美洲	北美洲	大洋洲
公元年 400	153	95	32	17	8	1	
1	252	170	43	26	12	1	
200	257	158	57	30	11	1	
600	208	134	33	24	16	1	
1000	253	152	43	39	18	1	
1200	400	258	66	48	26	2	
1340	442	238	90	80	32	2	
1400	375	201	65	68	39	2	
1500	461	245	84	87	42	3	
1600	578	338	111	113	13	3	
1700	680	433	125	107	12	3	
1750	791	502	163	106	16	2	2
1800	978	635	203	107	24	7	2
1850	1262	809	276	111	38	26	2
1900	1650	947	408	133	74	82	6
1950	2519	1398	547	221	167	172	12. 8
1955	2756	1542	575	247	191	187	14. 3
1960	2982	1674	601	277	209	204	15. 9
1965	3335	1899	634	314	250	219	17. 6
1970	3692	2143	656	357	285	232	19. 4
1975	4068	2397	675	408	322	243	21. 5
1980	4435	2632	692	470	361	256	22. 8
1985	4831	2887	706	542	401	269	24. 7
1990	5263	3168	721	622	441	283	26. 7
1995	5674	3430	727	707	481	299	28. 9
2000	6070	3680	728	796	520	316	31
2005	6454	3917	725	888	558	332	32. 9
2008	6707	4054	732	973	577	337	34. 3

表 5-6 世界及各大洲历史人口年均增长率(‰)

年份	世界	亚洲	欧洲	非洲	拉丁美洲	北美洲	大洋洲
—400—1	1.244	1.451	0.737	1.060	1.011		0.000
1—200	0.099	—0.368	1.416	0.719	—0.437		0.000
200—600	—0.529	—0.412	—1.366	—0.558	0.937		0.000
600—1000	0.490	0.315	0.662	1.214	0.294		0.000
1000—1200	2.290	2.645	2.142	1.038	1.839		3.466
1200—1340	0.713	—0.576	2.215	3.649	1.483		0.000
1340—1400	—2.740	—2.816	—5.424	—2.709	3.297		0.000
1400—1500	2.065	1.980	2.564	2.464	0.741		4.055
1500—1600	2.262	3.218	2.787	2.615	—11.727		0.000
1600—1700	1.625	2.477	1.188	—0.546	—0.800		0.000
1700—1750	3.024	2.957	5.309	—0.188	8.109		—8.109
1750—1800	4.244	4.700	4.389	0.188	8.109	25.055	0.000
1800—1850	5.099	4.843	6.144	0.734	9.191	26.244	0.000
1850—1900	5.362	3.150	7.817	3.616	13.330	22.972	21.972
1900—1950	8.462	7.790	5.864	10.156	16.279	14.816	15.154
1950—1955	17.984	19.608	9.984	22.245	26.856	16.723	22.163
1955—1960	15.763	16.427	8.845	22.926	18.012	17.402	21.212
1960—1965	22.376	25.222	10.691	25.075	35.825	14.190	20.316
1965—1970	20.339	24.176	6.822	25.669	26.206	11.533	19.475
1970—1975	19.397	22.402	5.710	26.706	24.412	9.265	20.556
1975—1980	17.275	18.705	4.975	28.293	22.865	10.423	11.742
1980—1985	17.105	18.495	4.006	28.507	21.017	9.907	16.009
1985—1990	17.130	18.577	4.205	27.535	19.017	10.147	15.572
1990—1995	15.039	15.892	1.657	25.618	17.364	10.999	15.836
1995—2000	13.493	14.070	0.275	23.714	15.592	11.060	14.029
2000—2005	12.268	12.483	—0.826	21.875	14.106	9.879	11.897
2005—2008	12.817	11.459	3.203	30.471	11.161	4.983	13.891

表 5-7　印度人口及其年均增长率

年份	1951	1961	1971	1981	1991	2001	2011
人口(百万)	361	439	548	683	846	1028	1210
年均增长率(‰)		19.562	22.178	22.022	21.402	19.485	16.301

5.4.2　世界历史人口演变规律

(1)人口年净增长率

纵观中外历史，在社会稳定发展时期，地区人口平均年净增长率可达到2%～3%。表 5-7 列出印度近 60 年人口增长数据，之所以以印度为例，是因为印度是人口大国但又没有强制执行计划生育政策，它的人口年均增长率是2%多一点。相对而言，非洲的年均增长率更高些，在 2.5%以上。应该说，非洲的数据更能代表纯自然的增长率。

记 x 为地区人口年净增长率，$y(t_0)$为第 t_0 年地区人口总数，$y(t)$为第 t 年地区人口总数。则

$$y(t)=y(t_0)(1+x)^{(t-t_0)}$$

假设 $y(t)=2y(t_0)$，则$(1+x)^t=2$，即 $t-t_0=\ln2/\ln(1+x)$。令 $x=2\%$，2.5%，3%，得 $t-t_0=23.45, 28.07, 35.00$(年)。因此在和平年代，大约平均每 25～30 年地区人口翻一番。

(2)人口增长率解释

是什么阻止了野兔在地球上繁殖过多？不是缺乏生育能力，而是另有原因，即天敌众多和食物不足。人类一般不会遭受天敌的威胁，影响最大的是战争和疾病。随着人类脱离野蛮状态的程度不断提高，抑制人口的因素愈来愈多地是对贫困的畏惧，而不是贫困本身。即使没有饿死的问题，很多人也会因为担心失去体面的生活而抑制生育。大多数人一旦不受外界的控制，就会尽早结婚，而且在大多数国家会多生子女。中等阶层当中的许多人还受改善生活境况的欲望的抑制，但这种欲望在劳动阶层中却很少出现或很少起作用。

在非常落后的社会状态下，如在中世纪的欧洲和目前非洲的许多地区，人口主要是靠饥饿来抑制的。在荒年，饥馑和随之而来的病疫，会夺去很多人的生命，而随后连续几年的好年景会使人口再度膨胀，遇到荒年又大批饿死。在较进步的社会状态下，即使在最贫穷的人当中，几乎也没有人仅仅能维持最低限度的生活。限制人口增长的因素，不是超量死亡，而是节制生育。生育可以用各种方法来节制。人们往往有其所习惯的生活，他们会察觉到，

如果家庭人口过多，他们的生活水平必然会降低，或无法使他们的子女保持目前的生活水平。挪威和瑞士的部分地区自觉地大规模节制生育时间，马尔萨斯曾仔细地收集了许多资料。在这两个国家，人口增长都很缓慢，抑制人口增长的，不是人口的大量死亡，而是出生人数较少。出生率和死亡率都非常低，平均寿命欧洲是最长的。人口中孩子所占的比例较低，壮年人所占的比例则比世界上任何其他国家都高。毫无疑问，低出生率不仅有助于消除产生贫困的主要根源，而且也有助于消除疾病的根源。

另一种情形，是由国家为人民深谋远虑，制定政策调控人口增长，如中国实行的计划生育政策。还有，在英国农村，19 世纪人口的增长非常有效地被住房难所抑制。按照当时的习惯，未婚的劳动者可以在雇主那里寄居和搭伙，已婚的劳动者则必须有房子。同时英国的济贫法规定，教区必须供养其失业的贫民，这使得地主们不愿意促成婚事。19 世纪末，战争和制造业对人力的巨大需求，致使人们把促进人口增长看成是爱国行为，大约与此同时，租地农场主长期受益于高涨的谷价，日益想过富人的生活，想让下等人住得远些，加上可以滥用济贫法获利的动机，他们逐步把劳动者赶入了农舍，地主们现在也不再拒绝允许盖房子了。

因此，总体上说，影响人口增长的主要因素是经济和文化。

①经济因素对人口自然增长的作用主要表现在它决定了人口的增殖条件和生存条件，通过改变人口的出生率和死亡率来影响人口的自然增长率。一般情况下，当人口数量不能满足经济发展对劳动力的需求时，人口自身的再生产必将会被刺激；当人口数量超越了经济发展所能提供的消费总数后，人口自身的再生产必将受到遏制。在现代生产力条件下，人口的自然增长率往往随着经济水平的提高而下降。经济因素对人口机械增长也有重要影响。通常情况下，经济发达或发展速度较快的地区，对人口具有一种吸引力和凝聚力，人口机械增长为正值；相反，经济落后或经济发展速度缓慢的地区，对人口会产生一种排斥力和离散力，人口机械增长一般为负值。

②文化因素更多地影响着人口的自然增长。随着科学文化水平的提高，人口自然增长率趋于下降，现代社会里这一趋势尤为明显，其表现主要在三个方面：一是由于人们受教育年限的延长，平均婚龄也会相应推延；二是科学文化水平愈发达，人们的生理知识、育儿知识、保健知识就愈丰富，促成婴儿死亡率降低；三是人们的科学文化水平愈高，就愈加注意自身及其后代各项素质的提高，少生优育，把有限的收入用于将子女培养成具有更高科学文化素质的现代人。

另外还有医疗卫生因素，医学的进步和医疗卫生事业的发展对人口出生

率和死亡率有着直接影响。首先，它使得因各种疾病致死的死亡率下降，从而降低人口死亡率，延长人口平均寿命；其次，它对控制生育和实行优生优育有着积极的作用。

非常明显，世界总人口或个别地区人口变化与中国人口变化类似，都基本上遵循马尔萨斯观点。在和平时期，人口按级数增长；随着粮食越来越紧张，人口自然增长率逐渐下降；社会出现动荡，并最终导致人口下降。更详细的原因不作描述，留作思考题给读者。

5.5 更精细的人口模型

5.5.1 人口发展方程

马尔萨斯模型是从宏观角度描述人口种群的变化规模，现在从每个年龄段的微观角度来考虑。如果我们知道今年每个年龄段数据特征：总数、男女比例、妇女生育率、存活率、死亡率、迁入迁出率，那么下一年的人口情况就基本预先知晓了。

设 $x(h,t)$，$\mu(h,t)$，$\omega(h,t)$，$\gamma(h,t)$，$v(h,t)$分别表示第 t 年年龄为 h 的总人口数、到下一年仍还活着的比例、育龄女性的比例、该年龄女性生育率（包括多胞胎和初生婴儿死亡的折算）、人口的迁入和迁出数，则

$$x(h,t+1)=\mu(h,t)x(h,t)+v(h,t),\ (h\geqslant 1)$$

$$x(0,t+1)=\sum^{h}\gamma(h,t)w(h,t)x(h,t)+v(0,t)$$

其中 $h=0$ 特指新生儿。以上模型通常叫做人口发展方程。

由于详细人口数据非常多，本节摘取国务院全国1%人口抽样调查领导小组办公室和国家统计局人口和就业统计司编制的《2005年全国1%人口抽样调查资料》中的部分数据。这次调查的标准时间为2005年11月1日零时，调查对象是在被抽中的调查小区内具有中华人民共和国国籍并于2005年10月31日晚居住在本调查小区以及户口在本户、2005年10月31日晚未居住在本户的全部人口；共调查登记的常住人口1699万，占全国总人口的1.325%，数据见表5-8。

表 5-8　2005 年全国 1%人口抽样调查部分数据

年龄	人口数			平均育龄妇女人数	出生人数	生育率(%)	上一年死亡率(%)		
	合计	男	女				合计	男	女
20	196962	97506	99456	99451	6735	67.72	0.86	1.28	0.47
21	191823	91680	100143	101725	11016	108.30	0.86	1.21	0.54
22	198985	95182	103803	113155	13947	123.25	1.00	1.44	0.61
23	235835	112784	123051	119600	15891	132.87	0.77	1.13	0.44
24	213119	102776	110344	106783	14300	133.91	0.88	1.27	0.51
25	210881	100577	110304	115252	13794	119.69	0.94	1.20	0.70
26	223647	108750	114897	113871	11978	105.19	0.84	1.20	0.50
27	222032	108001	114031	110544	10037	90.79	0.96	1.40	0.53
28	214130	104788	109343	115952	9082	78.32	1.00	1.50	0.52
29	239600	117121	122480	126327	8477	67.10	1.22	1.72	0.74
30	251922	124107	127815	133869	7660	57.22	1.05	1.49	0.63

此处忽略了中国公民移居国外和外国人定居中国的情况。虽然死亡率是上一年 2004 年 11 月至 2005 年 10 月的数据，但我们完全可以近似地假设 2005 年 11 月至 2006 年 10 月的每个年龄的死亡率与上一年度一样。这样 2006 年 10 月底 21 岁男性人口数＝97506×(1—1.28/100)≈96258。同样地假设这些 20～30 岁女性生育率与上一年度一样，则 99456 个 20 岁女性将生育 99456×67.72/100≈67352 个婴儿，到 2006 年 10 月底这些 20～30 岁女性共生育婴儿＝67352＋…＋73136＝1212372 个婴儿。结果见表 5-9。

表 5-9　预测 2006 年 10 月底被调查家庭人口数和生育婴儿数

年龄	人口数			生育婴儿	年龄	人口数			生育婴儿
	合计	男	女			合计	男	女	
					26	208902	99370	109532	132023
21	195246	96258	98989	67352	27	221768	107445	114323	120860
22	190173	90571	99602	108455	28	219916	106489	113427	103529
23	196981	93811	103170	127937	29	211991	103216	108774	85637
24	234019	111510	122510	163498	30	236680	115107	121574	82184
25	211252	101471	109781	147762	31	249268	122258	127010	73136

5.5.2 变系数问题

上一小节后一年死亡率、生育率都直接采用上一年数据，就是假设死亡率、生育率等都不变的。但实际情况这些参数应该是变化的，简称为变系数模型。表 5-10 是我国 1978—2010 年总人口出生率、死亡率数据。

表 5-10 我国历年人口出生率和死亡率(%)

年份	出生率	死亡率	年份	出生率	死亡率	年份	出生率	死亡率
1978	18.25	6.25	1990	21.06	6.67	2001	13.38	6.43
1980	18.21	6.34	1991	19.68	6.70	2002	12.86	6.41
1981	20.91	6.36	1992	18.24	6.64	2003	12.41	6.40
1982	22.28	6.60	1993	18.09	6.64	2004	12.29	6.42
1983	20.19	6.90	1994	17.70	6.49	2005	12.40	6.51
1984	19.90	6.82	1995	17.12	6.57	2006	12.09	6.81
1985	21.04	6.78	1996	16.98	6.56	2007	12.10	6.93
1986	22.43	6.86	1997	16.57	6.51	2008	12.14	7.06
1987	23.33	6.72	1998	15.64	6.50	2009	11.95	7.08
1988	22.37	6.64	1999	14.64	6.46	2010	11.90	7.11
1989	21.58	6.54	2000	14.03	6.45			

注：本表数据来自 2011 年《中国统计年鉴》。

1978—1987 年我国出生率波动上升，1987 年后的出生率持续缓慢下降；死亡率的波动幅度很小，基本上多数年份维持在 6.50%多点。图 5-4 标出了我国历年出生率的散点图和趋势线，直观上看，21 世纪 10 年代的出生率估计在 7%左右。

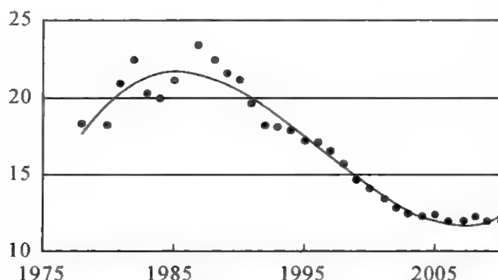


图 5-4 我国历年出生率散点图

对人口发展方程中的每个年龄段的男女比例、妇女生育率、存活率、死亡率、迁入迁出率都要先找出其变化趋势，再设法找出其变化的数学公式（通常称为经验回归方程），并估计某时刻的估计值，然后计算所需要的值。

思考题

1. 请自学回归分析方法，分别对本章提到的中国和世界历史人口数据计算回归方程并讨论其显著性。

2. 据对古代粮食生产史料初步研究估计，战国前期亩产粟 205.8 市斤，战国后期亩产粟约 247 市斤，汉朝亩产粟 281 市斤，唐代的亩产比汉代增长了 1/4 还多；两宋南方水稻亩产比唐代增长了 12.5%，但北方亩产只有 309 斤，比唐低；元代北方旱地的粮食亩产已经恢复到汉代的水平，但南方略低于唐；明清两朝单季稻谷亩产一直在 300 多斤水平。请查阅更多史料，研究中国历朝人口峰值与粮食生产的关系。（请参阅吴慧著：《中国历代粮食亩产研究》，农业出版社 1985 年版）。

3. 60 年执政规律。一个朝廷从建立到国家一统，随后休养生息，约 60 年后民富国强；但官员日渐腐败，财富逐渐往权贵们手中集中，社会矛盾加剧，这时需要有人出来肃整朝纲，然后皇朝才会朝着盛世挺进，这就是所谓的 60 年规律。请查阅中国历史文献资料，找两个朝代进行对比说明。

4. 请查阅日本、印度或其他国家历史资料，分析外国是否也有“60 年执政规律”。

5. 根据表 5-10 数据（也可利用其他文献如国家统计局公布的中国历史人口数据），计算 1978—2010 年我国人口自然增长率，画出死亡率和自然增长率的散点图，指出其随年份变动的趋势，利用 Excel 或其他统计软件计算出它们的经验回归方程。

6. 将人口发展方程写成矩阵形式，编写用发展方程估算人口数的算法。

7. 查阅人口普查或人口调查数据，得出某年中国全部年龄男女数、生育数、死亡数，利用人口发展方程估算 10 年后每个年龄人口数。

第6章 饮酒、吃药和打针——动力学模型

在许多实际问题的研究中，要直接导出变量之间的函数关系较为困难，但要导出包含未知函数的导数或微分的关系式却较为容易，此时即可用建立微分方程模型的方法来研究实际问题。本章介绍一些饮酒、吃药和打针、传染病等相关的微分方程问题。

6.1 饮酒驾车

6.1.1 饮酒驾车

据报道，2003 年全国道路交通事故死亡人数为 10.4372 万，其中因饮酒驾车造成的事故占有相当的比例。

针对这种情况，国家质量监督检验检疫局于 2004 年 5 月 31 日发布了新的《车辆驾驶人员血液、呼气酒精含量阈值与检验》国家标准。新标准规定，车辆驾驶人员血液中的酒精含量大于或等于 20 毫克/百毫升、小于 80 毫克/百毫升为饮酒驾车(原标准是小于 100 毫克/百毫升)，血液中的酒精含量大于或等于 80 毫克/百毫升为醉酒驾车(原标准是大于或等于 100 毫克/百毫升)。

大李在中午 12 点喝了一瓶啤酒，下午 6 点检查时符合新的驾车标准，紧接着他在吃晚饭时又喝了一瓶啤酒，为了保险起见，他待到凌晨 2 点才驾车回家，检查时被认定为饮酒驾车，这让他既懊恼又困惑，为什么喝同样多的酒，两次检查结果会不一样呢?

请你参考下面给出的数据(或自己收集资料)建立饮酒后血液中酒精含量的数学模型，并讨论以下问题。

(1)对大李碰到的情况作出解释。

(2)在喝了 3 瓶啤酒或者半斤低度白酒后多长时间内驾车就会违反上述标准，在以下情况下回答：

酒是在很短时间内喝的；

酒是在较长一段时间(比如 2 小时)内喝的。

(3)怎样估计血液中的酒精含量在什么时间最高。

(4)根据你的模型论证:如果天天喝酒,是否还能开车?

(5)根据你做的模型并结合新的国家标准写一篇短文,给想喝一点酒的司机如何驾车提出忠告。

参考数据:

(1)人的体液占人的体重的65%~70%,其中血液只占体重的7%左右;而药物(包括酒精)在血液中的含量与在体液中的含量大体是一样的。

(2)体重约70 kg的某人在短时间内喝下2瓶啤酒后,隔一定时间测量他血液中酒精含量(mg/100mL),得到数据如表6-1。

表6-1 体重约70 kg的某人在短时间内喝下2瓶啤酒后血液中酒精含量(mg/100mL)

时间(h)	0.25	0.5	0.75	1	1.5	2	2.5	3	3.5	4	4.5	5
酒精含量	30	68	75	82	82	77	68	68	58	51	50	41
时间(h)	6	7	8	9	10	11	12	13	14	15	16	
酒精含量	38	35	28	25	18	15	12	10	7	7	4	

图6-1是表6-1对应的散点图,血液中酒精含量随时间推移的变化规律明显,为什么是先快速上升后缓慢下降呢?利用医学知识可以解释这一现象,酒精入肚后经胃、肠以较快速度被吸收到血液中,同时肝脏持续但相对较慢地分解酒精,这就促成了图6-1的单峰偏态曲线。

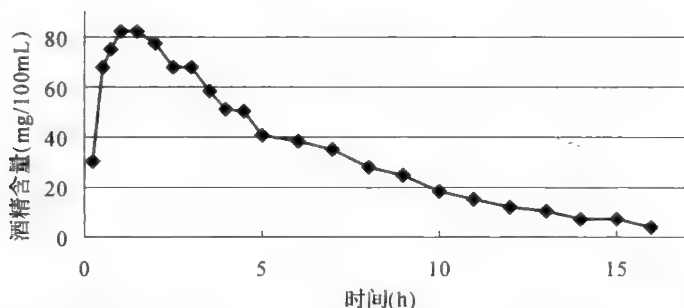


图6-1 体重约70 kg的某人在短时间内喝下2瓶啤酒后血液中酒精含量折线图

6.1.2 生活常识

(1)食物在肚子里的行程

有位叫张莉的初一学生写了篇非常生动形象的科普短文《食物旅行记》,摘抄如下。

食物从我们的口腔大门进入,牙齿像一排锋利的刀,把它切碎。舌头像

一个巨大的搅拌机，在多次搅拌下，食物和唾液混合后，食物中的淀粉被分解成麦芽糖。并且变成湿润的食物团被咽了下去。

接着，食物到了食管。食管像一个又细又长的滑道，缓慢蠕动促进食物进入胃。

胃像一个上下开口的大口袋，上面连接着食管，下面连接着小肠。食物中的水分和酒精被胃吸收。它分泌出酸性胃液，把食物变成稀糊状。肝脏是人体最大的消化腺，它可以分泌胆汁，帮助消化。

接下来，食物就到了小肠。小肠有 5～7 米长，它的表面布满了肉眼看不见的绒毛，像一条卷起来的毛巾，这样的结构增加了小肠与食物接触的面积，有利于吸收营养物质。经胆汁乳化的脂肪微粒和麦芽糖等物质与未消化的食物，在小肠中被三种消化液（胰液、肠液和胃液）分解成简单的物质。淀粉被分解成葡萄糖，脂肪被分解成脂肪酸和甘油，蛋白质被分解成氨基酸。这些简单的物质被小肠吸收进入血液循环系统。

紧接着，食物来到了又粗又短的大肠。大肠主要吸收食物中剩余的水、无机盐、矿物质和维生素。

食物经过了漫长的旅程，营养都被人体吸收了，变成了臭烘烘的粪便，从肛门排出体外。

(2)饮酒生理学知识

摄入体内的酒精（乙醇）除极少量经呼吸和尿排泄外，95%以上在体内分解代谢，而肝脏是乙醇代谢的重要器官。在周围组织内进一步氧化为二氧化碳和水，其余在肝内进入糖和（或）脂肪池，或进入三羧酸循环而氧化分解。乙醇和乙醛都可以使人出现头晕、脸红、心跳过速甚至神志不清等酒精中毒现象，但乙醛的作用比乙醇更大。酒精在肝内的代谢带来多种后果：刺激脂肪的合成，消耗大量的氧，给肝脏造成缺氧状态，干扰肝细胞 ATP 的产生，影响蛋白质的合成，造成直接损伤，出现肝功能障碍。

酒的化学成分是乙醇，在消化道内不需要消化即可吸收，吸收快而且完全。一般在胃中吸收 20%，其余 80%被十二指肠和大肠吸收。胃内有无食物、胃壁的功能状况、饮料含酒精的多少以及饮酒习惯均可影响酒精的吸收。空腹饮酒时，15 分钟吸收 50%左右，半小时吸收 60%～90%，2～3 小时吸收 100%。酒精还能通过皮肤和呼吸道进入体内。酒对人体的作用与其浓度和吸收速度成正比，即浓度越高，吸收速度越快，作用也越明显。

进入人体内的酒，约 10%经由呼吸道或者以尿液和汗液的形式排出。因此，饮酒者都是“一身酒气”，也可用呼吸测酒器检测出来，其余 90%经由肝脏代谢。乙醇首先被氧化成乙醛，脱氢后转化为乙酸，最后氧化成二氧化碳和

水排出体外，同时放出大量的热能。但乙醇的氧化，并不受血液中酒精浓度高低的影响，也不按机体的需要进行，它只按其固定的规律进行，即肝脏以每小时 10 毫升的速度将酒精分解成水、二氧化碳和糖，直至消化完为止。

许多因素会影响乙醇在肠胃道的吸收，空胃饮酒所引起的酒精毒害最显著。食物可以减慢乙醇的吸收速率，还可延缓血液中酒精高峰期的到达，除此之外，食物的成分及量都会直接影响乙醇在肠胃道的吸收，例如，可溶解的碳水化合物对于延缓乙醇吸收的作用大于蛋白质及脂肪。

大量摄入乙醇对肠道会有直接的毒害作用，尤其是小肠。酒精中毒者有痉挛性的腹部疼痛，特别是在狂饮后。这是因为乙醇引起双糖的缺乏而造成乳糖不耐及小肠吸收水分和电解质缺损的结果。

人饮用啤酒后，对小肠做的生检结果显示，尽管啤酒有高含量的麦芽糖，然而小肠内麦芽糖和蔗糖的活性却仍然减少；禁酒两周后，这两种双糖才又增加。

乙醇在体内有 90%以上会被氧化成水和二氧化碳，在氧化过程中，每克乙醇产生 7 大卡的热量。虽然某些酒精饮料中含有少量糖类、微量元素或维生素，但主要成分仍是乙醇，因此除了热量外，酒的营养价值极低。代谢后产生的热量如未被使用，则剩余的部分会以脂肪形态贮存；少部分未代谢的乙醇则由尿液、呼出的气体、汗液及乳汁排出。

肝脏是乙醇氧化的主要部位，人体内只有 10%到 15%的乙醇在其他组织氧化。在肝脏中乙醇先被酒精去氢酶(ADH)氧化成乙醛，乙醛再经去氢氧化成醋酸。这两种去氢将其各别受质上的氢离子转移到氧化形的烟酰胺腺嘌呤双核酸(NAD^+)上。酒精代谢的速率限制步骤是：乙醇被 ADH 氧化成乙醛的过程。酒精经上述两种代谢时，会使还原形的 NADH 对 NAD^+ 的比例增加，而改变有机体的氧化还原状态；这种改变会使糖新生成受损及产生酒精性酮中毒。醋酸是酒精代谢最后产物，在肝脏中可转变为醋酸辅 A 而进入醋酸代谢的正常途径。

6.2 房室模型

在药物动力学中通常用房室模拟人体，只要体内某些部位接受或消除药物的速率相似，即可归入一个房室。房室模型仅是进行药物动力学分析的一种抽象概念，并不一定代表某一特定解剖部位。

房室具有以下特征：它由考察对象均匀分布而成(实际上考察对象一般并非均匀分布，这里采用了一种简化方法)；房室中考察对象的数量或浓度

（密度）的变化率与外部环境有关，这种关系被称为“交换”且交换满足总量守衡。图 6-2 是一单房室系统。

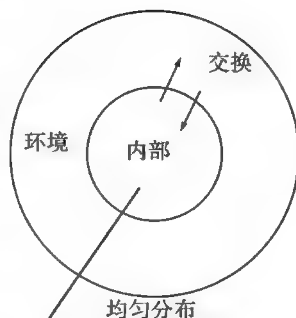


图 6-2 单房室系统

把机体划分为一个或多个独立单元，可对药物在体内吸收、分布、消除的特性作出模式图，以建立数学模型，揭示其动态变化规律。

（1）假设机体给药后，药物立即在全身各部位达到动态平衡，这时把整个机体视为一个房室，称为一室模型或单室模型。

（2）假设药物进入机体后，瞬时就可在血液供应丰富的组织（如血液、肝、肾等）分布达到动态平衡，然后再在血液供应较少或血流较慢的组织（如脂肪、皮肤、骨骼等）分布达到动态平衡，此时可把这些组织分别称为中央室和周边室，即二室模型。

多数情况下，二室模型能够准确地反映药物的体内过程特征，但一房室模型虽然准确性稍差，却比较简单，便于理解、推广、应用，且有些药物用单室模型处理已能满足要求，所以其重要性并不亚于二室模型。

6.2.1 单房室模型

（1）模型假设

为了便于问题描述和处理，作如下假设：

体内药物在任一时刻都是均匀分布的，机体的容积（即血液体积或药物分布容积）在动态过程中保持不变；

药物的分解与排除（输出）速率与药物当前的浓度成正比；

只有机体与环境有药物交换，即药物从体外进入机体，最后又从机体排出体外（同分解与排除的数量相比，药物的吸收可以忽略）。

（2）模型建立

设 t 时刻体内药物的总量为 $x(t)$，初始药量 $x(0)=x_0$，房室容积为 V，给药速率 $\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)_{\text{入}}=f(t)$，由假设，药物的分解与排除（输出）速率同药物当前的浓度成正比，即 $\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)_{\text{出}}=kx$，其中 k 为药物排除速率系数，则

$$\frac{\mathrm{d}x}{\mathrm{d}t}=\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)_{\text{入}}-\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)_{\text{出}}$$

满足初始条件的药量的微分方程：

$$\begin{cases}\dfrac{dx}{dt}+kx=f(t)\\ x(0)=x_0\end{cases}$$

$x(t)$与血药浓度 $c(t)$ 及房室容积 V 之间显然存在如下关系：$x(t)=V\,c(t)$。回代可得到血药浓度的微分方程：

$$\begin{cases}\dfrac{dc}{dt}+kc=\dfrac{f(t)}{V}\\ c(0)=c_0\end{cases}$$

其中初始浓度为 c_0。建立的单房室模型如图 6-3。

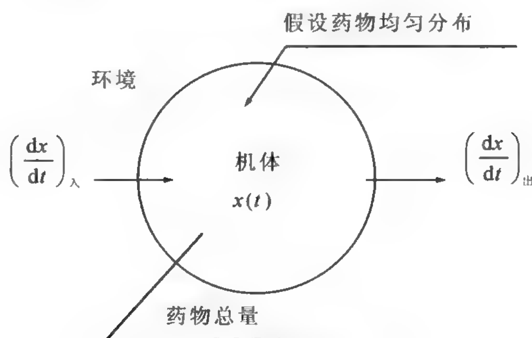


图 6-3 单房室模型

药物的输入规律与给药的方式有关。下面我们来研究在几种常见的给药方式下体内药体的变化规律。

①情况 1：快速静脉注射。

在快速静脉注射时，总量为 D 的药物在瞬间被注入体内，则我们可以近似地将系统看成初始总量为 D，浓度为 D/V，只输出不输入的房室，即 $f(t)=0$。如图 6-4。

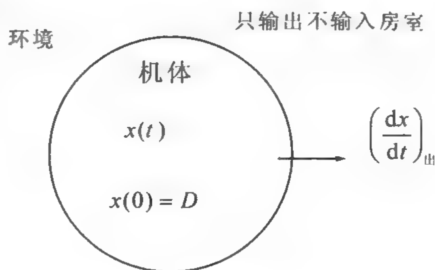


图 6-4 快速静脉注射

系统可看成近似地满足微分方程：

$$\begin{cases}\dfrac{dx}{dt}+kx=0\\ x(0)=D\end{cases}$$

分离变量，得

$$\frac{1}{x}dx=-k\,dt$$

两端积分，并整理得通解为

$$x(t)=Ce^{-kt}$$

代入初始条件 $x(0)=D$，得 $C=D$。故原方程解为

$$x(t)=De^{-kt}$$

药物浓度：

$$c(t)=\frac{D}{V}e^{-kt}$$

与放射性物质类似，医学上将血浆药物浓度衰减一半所需的时间称为药物的血浆半衰期：$t_{1/2}=\ln2/k$。

②情况 2：恒速静脉注射。

药物以恒速点滴方式进入体内，即 $f(t)=\left(\frac{dx}{dt}\right)_{入}=K_0$，如图 6-5。

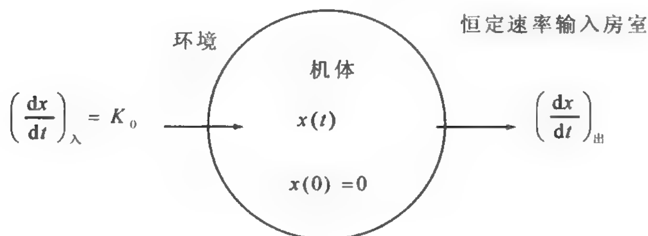


图 6-5　恒速静脉注射

满足初始条件的药量的微分方程：

$$\begin{cases}\dfrac{dx}{dt}+kx=K_0\\ x(0)=0\end{cases}$$

这是一个一阶常系数非齐次线性方程，其对应的齐次方程的通解为：

$$x(t)=Ce^{-kt}$$

用常数变易法，把 C 换成 u，即令 $x(t)=ue^{-kt}$，那么

$$\frac{dx}{dt}=u'e^{-kt}-kue^{-kt}$$

代入所给非齐次方程，得

$$u'=K_0e^{kt}$$

两端积分得

$$u=\frac{K_0}{k}e^{kt}+C$$

得所求方程的通解为

$$x(t)=\frac{K_0}{k}\left(1+\frac{kC}{K_0}e^{-kt}\right)$$

代入初始条件 $x(0)=0$，得 $C=\frac{K_0}{k}$，所以非齐次方程的解为

$$x(t)=\frac{K_0}{k}(1-e^{-kt})$$

药物浓度：

$$c(t)=\frac{K_0}{Vk}(1-e^{-kt})$$

易见：

$$\lim_{t\to+\infty}c(t)=\frac{K_0}{Vk}$$

对于多次点滴，设点滴时间为 T_1，两次点滴之间的间隔时间为 T_2，则在第一次点滴结束时病人体内的药物浓度可由上式得出，其后 T_2 时间内为情况 1，故

第一次：

$$c(t)=\frac{K_0}{Vk}(1-e^{-kT_1})(0\leqslant t\leqslant T_1)$$

第二次：

$$c(t)=\frac{K_0}{Vk}(1-e^{-kT_1})e^{-k(t-T_1)}(T_1\leqslant t\leqslant T_1+T_2)$$

类似可讨论以后各次点滴时的情况，区别只在初值上的不同。第二次点滴起，患者体内的初始药物浓度不为零.

③情况 3：口服药或肌肉注射。

口服药或肌肉注射时，药物的吸收方式与点滴时不同，药物虽然瞬间进入了体内，但它一般都集中于身体的某一部位，靠其表面与肌体接触而逐步被吸收，如图 6-6。设药物被吸收的速率与存量药物的数量成正比，记比例系数为 k_1，即若记 t 时刻残留药物量为 $y(t)$，则 $y(t)$ 满足：

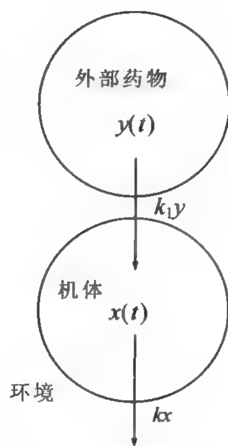


图 6-6　口服药或肌注

$$\begin{cases}\dfrac{dy}{dt}=-k_1y\\ y(0)=D\end{cases}$$

其中 D 为口服或肌注药物总量，解得

$$y(t)=De^{-k_1t}$$

所以药量微分方程：

$$\begin{cases}\dfrac{dx}{dt}+kx=k_1De^{-k_1t}\\ x(0)=D\end{cases}$$

解得：

$$x(t)=\frac{K_1 D}{k_1-k}(e^{-kt}-e^{-k_1 t})$$

药物浓度：

$$c(t)=\frac{k_1 D}{V(k_1-k)}(e^{-kt}-e^{-k_1 t})$$

(3)模型分析

图 6-7 给出了上述三种情况下体内血药浓度的变化曲线。容易看出，快速静脉注射能使血药浓度立即达到峰值，常用于急救等紧急情况；口服、肌注与点滴也有一定的差异，主要表现在血药浓度的峰值出现在不同的时刻，血药的有效浓度保持时间也不尽相同。

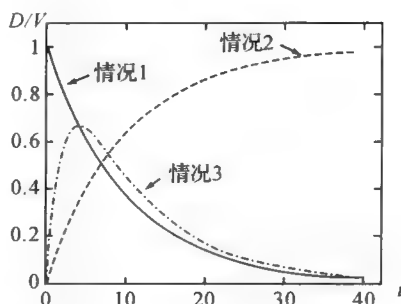


图 6-7　血药浓度变化曲线

我们已求得三种常见给药方式下的血药浓度 $c(t)$，当然也容易求得血药浓度的峰值及出现峰值的时间，因而，也不难根据不同疾病的治疗要求找出最佳治疗方案。

上述研究是将机体看成一个均匀分布的同质单元，故称单房室模型，但机体事实上并不是这样。药物进入血液、通过血液循环被带到身体的各个部位，又通过交换进入各个器官。因此，要建立更接近实际情况的数学模型就必须正视机体部位之间的差异及相互之间的关联关系，这就需要运用多房室系统模型。

6.2.3　二房室模型

将一个机体分为若干个房室，假定每个房室内药物呈均匀分布，即血药浓度是常数，不同房室之间，按一定规律进行药物转移。

一个机体要分为几个房室可以依据不同药物的吸收、分布、排除的具体情况确定或根据所要求的精确度而决定。例如二室模型将机体分为血液丰富的中心室(如心、肺、肝、肾等)和血液贫乏的周边态(如肌肉组织等)。

当差异较大的部分较多时，可以类似建立多房室系统，即 N 房室系统。图 6-8 为两房室系统。

(1)模型假设

为了便于问题描述和处理，作如下假设：

机体分为中心室(Ⅰ类)和周边室(Ⅱ类)，并假定两个室的容积(即血液容积或药物分布的容积)在过程中不变。在每个房屋内血液浓度均匀分布，

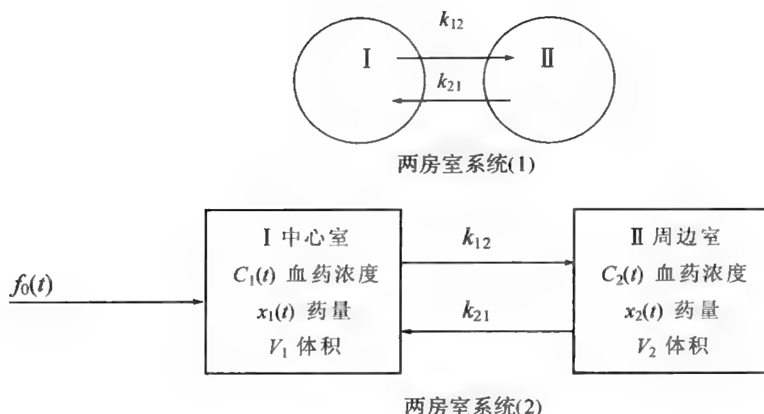


图 6-8　两房室系统

即为常数。

药物从一室向另一室的转移速率及向体外的排除速率与该室的血药浓度成正比：血药浓度大，则转移速率和排除速率快；血药浓度小，则转移速率和排除速率慢。

只有中心室与体外有药物交换，即从体外进入中心室，又从中心室排出体外，与药物的转移与排除的数量相比，药物的吸收可以忽略。

(2)模型符号

$C_1(t)$，$x_1(t)$，V_1 表示第Ⅰ室血药浓度、药量、容积；

$C_2(t)$，$x_2(t)$，V_2 表示第Ⅱ室血药浓度、药量、容积；

k_{12} 为药物由Ⅰ室转移到Ⅱ室的转移速率系数；

k_{21} 为药物由Ⅱ室转移到Ⅰ室的转移速率系数；

k_{13} 为药物由Ⅰ室转移到体外的排除速率系数；

$f_0(t)$ 表示给药速率；

D_0 表示给药剂量($t=0$ 时的初始值)。

以上一级速率系数 k_{ij} 为常数时的房室模型，称乳突状模型。

(3)模型建立

药量：$x_1(t)$，$x_2(t)$ 满足的微分方程为：

$$\begin{cases}\dot{x}_1(t)=-k_{12}x_1(t)-k_{13}x_1(t)+k_{21}x_2(t)+f_0(t)\\ \dot{x}_2(t)=k_{12}x_1(t)-k_{21}x_2(t)\end{cases}$$

由 $x_1(t)=V_1C_1(t)$ 和 $x_2(t)=V_2C_2(t)$，代入上式得：

$$\begin{cases}\dot{C}_1(t)=-(k_{12}+k_{13})C_1(t)+k_{21}\dfrac{V_2}{V_1}C_2(t)+f_0(t)/V_1\\ \dot{C}_2(t)=\dfrac{V_1}{V_2}k_{12}C_1(t)-k_{21}C_2(t)\end{cases}$$

此为一阶带系数线性非齐次常微分方程组，去掉 $f_0(t)/V_1$ 为其对应的齐次方程组。通解为：

$$\begin{cases}\overline{C}_1(t)=A_1e^{-\alpha t}+B_1e^{-\beta t}\\ \overline{C}_2(t)=A_2e^{-\alpha t}+B_2e^{-\beta t}\end{cases}$$

其中 α,β 由方程组 $\begin{cases}\alpha+\beta=k_{12}+k_{21}+k_{13}\\ \alpha\beta=k_{21}k_{13}\end{cases}$ 确定。

(4)模型求解

为求出非齐方程的通解，需依非齐次项 $f_0(t)$ 和初始条件来确定，为此需考虑以下几种不同的给药方式：(1)快速静脉注射(静脉注射)；(2)恒速静脉滴注(吊针)；(3)口服或肌肉注射(肌肉注射)。

①快速静脉注射。

在 $t=0$ 时，将剂量 D_0 的药物输入中心室，于是有：

$$f_0(t)=0,\quad C_1(0)=\frac{D_0}{V_1},\quad C_2(0)=0$$

于是原方程为齐次方程，其解为：

$$\begin{cases}C_1(t)=Ae^{-\alpha t}+Be^{-\beta t}\\ C_2(t)=\dfrac{D_0k_{12}}{V_2(\beta-\alpha)}(e^{-\alpha t}-e^{-\beta t})\end{cases}$$

其中：

$$A=\frac{D_0(k_{21}-\alpha)}{V_1(\beta-\alpha)},\quad B=\frac{D_0(\beta-k_{21})}{V_1(\beta-\alpha)}$$

α,β 由 $\begin{cases}\alpha+\beta=k_{12}+k_{21}+k_{13}\\ \alpha\beta=k_{21}k_{13}\end{cases}$ 确定。

特别地，当 $t\to\infty$ 时，$C_1(t)\to0$，$C_2(t)\to0$。

② 恒速静脉滴注。

当静脉滴注的速率为 k_0 时，$f_0(t)$ 和初始条件为：

$$f_0(t)=k_0,\ C_1(0)=0,\ C_2(0)=0$$

则：

$$\dot{C}_1(t)=-(k_{12}+k_{13})C_1(t)=\frac{V_2}{V_1}k_{21}C_2(t)+\frac{f_0(t)}{V_1}$$

$$\dot{C}_2(t)=\frac{V_1}{V_2}k_{12}C_1(t)-k_{21}C_2(t)$$

其特解：

$$\begin{cases} C_1(t)=A_1e^{-\alpha t}+B_1e^{-\beta t}+\dfrac{k_0}{k_{13}V_1} \\ C_2(t)=A_2e^{-\alpha t}+B_2e^{-\beta t}+\dfrac{k_{12}k_0}{k_{21}k_{13}V_2} \\ A_2=\dfrac{V_1(k_{12}+k_{13}-\alpha)}{k_{21}V_2}A_1,B_2=\dfrac{V_1(k_{12}+k_{13}-\beta)}{k_{21}V_2}B_1 \end{cases}$$

其中常数 A_1,B_1 由初始条件 $C_1(0)=C_2(0)$确定。

由上式中 $C_1(t),C_2(t)$的表达式可知：

当 $t\to\infty$时，血药浓度 $C_1(t)$将趋向于$\dfrac{k_0}{k_{13}V_1}$，$C_2(t)$将趋向于$\dfrac{k_{12}k_0}{k_{21}k_{13}V_2}$；

实际上不可能 $t\to\infty$，当 $t=T$ 时停止滴注，则 $C_1(t),C_2(t)$在 $t>T$ 时将按指数规律衰减并趋向于 0。

③口服或肌肉注射。

这种给药方式相当于在药物进入中心室之前，先有一个将药物吸收入血液的过程，因而可简化为有一个吸收室，即成三房室系统。

$x_0(t)$：吸收室的药量(t 时刻)；

k_{01}：药物由吸收室进入中心室的转移速率系数；

(给药方式剂量)给药速率 $f_0(t)=k_{01}\,x_0(t)$；

D_0：$t=0$ 时给药量，即 $D_0=x(0)$。

于是，$x_0(t)$满足：

$$\begin{cases} \dot{x}_0(t)=-f_0(t)=-k_{01}x_0(t) \\ \dot{x}_0(0)=D_0 \end{cases}$$

有 $x_0(t)=D_0\exp\{-k_{01}\,t\}$。因此给药速率为：

$$f_0(t)=k_{01}x_0(t)=D_0k_{01}\exp\{-k_{01}\,t\}$$

于是，药物由中心室向周边室传送的血药浓度由下列方程组可确定：

$$\begin{cases} \dot{C}_1(t)=-(k_{12}+k_{13})C_1(t)+\dfrac{V_2}{V_1}k_{21}C_2(t)=\dfrac{f_0(t)}{V_1} \\ \dot{C}_2(t)=\dfrac{V_1}{V_2}K_{12}C_1(t)-k_{21}C_2(t) \end{cases}$$

其初始条件：$C_1(0)=C_2(0)=0$，且 $f_0(t)=D_0\,k_{01}\exp\{-k_{01}\,t\}$。

由上述非齐次方程组的通解可得：

$$C_1(t)=Ae^{-\alpha t}+Be^{-\beta t}+Ee^{-k_{01}t}$$

其中：α,β 由方程 $\alpha+\beta=k_{12}+k_{21}+k_{13}$ 和 $\alpha\beta=k_{21}\,k_{13}$ 确定；$k_{01}\neq\alpha,\beta$；A,B,E 由初始条件 $C_1(0)=C_2(0)=0$ 确定。

(5)模型校正及讨论

要计算血药浓度 $C_1(t),C_2(t)$ 的变化规律，需要已知参数：血药转移速率

系数 k_{01}，k_{12}，k_{21}，k_{13}，房室容积 V_1，V_2，给药量 D_0 等。

然而在实际应用中正好相反，即通过对 $C_1(t)$，$C_2(t)$的测量来确定药理学和临床医学最重要的参数，如转移速率系数 k_{01}，k_{12}，k_{21}，k_{13}，特别是排除速率系数 k_{13}的解。此即是微分方程问题的反解，即参数讨论问题。由于内容较多，此处省略。

6.3 饮酒驾车的数学模型

6.3.1 模型假设

(1)模型假设

假设酒精在胃肠道中的吸收是一级动力学过程，即胃肠道中酒精吸收率与胃肠道中酒精含量成正比；

忽略从饮酒到酒精开始吸收的时间延迟；

忽略食物摄入对酒精吸收速率造成的影响；

由于呼吸和排尿对体液中的酒精总量影响很小，因此不考虑呼吸、尿液对酒精总量的影响；

酒精在血液和体液中的分布相同。

(2)主要变量及符号假设

n：喝酒的瓶数；

m：每瓶啤酒中的酒精含量(mg)；

f：酒精的吸收率；

r：多次饮酒时喝酒的次数；

V：人体体液的体积(百毫升)；

k_0：酒精的吸收常数；

k：酒精的消除常数；

A_i：第 i 次喝酒前血液中酒精含量；

W_i：第 i 次喝酒前胃肠道中酒精含量。

6.3.2 模型建立

(1)问题分析

通过对问题的分析，可以认为大李血液中的酒精含量超标主要是由两次饮酒前血液中酒精初始浓度不同、喝酒次数和喝酒时间长短不同造成的。当酒是在短时间内喝下，胃肠道内酒精总量为 fnm，可用一级动力学过程建立

微分方程处理;当酒是分多次等量等时间间隔在较长时间内喝下,就形成一个多剂量给药问题,运用叠加的方法可解决此问题。酒精在血液中的浓度必存在一个峰值,我们使用药动力学对峰值的大小和取峰值的时间进行了计算。最后讨论的是每天定时等量喝一次酒的情况,主要考虑了饮酒的时间和饮酒量的变化对血液中酒精浓度的影响。

酒精是一种常被人们饮用的化学物质,它必须通过胃肠道的吸收才能进入血循环,然后才能在体内分布和逐渐消除。也就是说,个体饮酒后,酒精首先迅速进入吸收室(如胃肠道),在该室逐渐吸收,然后同时将吸收到的酒精逐渐转运至中心室(如血液、肝脏),并在中心室内得到降解和消除。可用图6-9来直观地表示酒精在体内吸收和排除的过程。

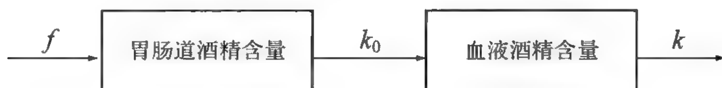


图 6-9　饮酒模型框

①一次性快速喝下 n 瓶酒。设 $y(t)$ 为 t 时刻血液中酒精含量,$x(t)$ 为 t 时刻胃肠道中酒精含量。快速喝酒时,血液中酒精浓度达最大值的时间为 t_{p1}。可列出微分方程组

$$\begin{cases} x'(t) = -k_0 x(t) \\ y'(t) = k_0 x(t) - k y(t) \\ x(0) = f \cdot n \cdot m \\ y(0) = 0 \end{cases}$$

解得 t 时刻血液中酒精浓度为

$$Y(t) = \frac{y(t)}{V} = \frac{f \cdot n \cdot m \cdot k_0}{V \cdot (k_0 - k)} (\mathrm{e}^{-kt} - \mathrm{e}^{-k_0 t}) (\mathrm{mg}/100\ \mathrm{mL})$$

fm/V,k 与 k_0 可用题中所给的实验数据拟合求得。

下面我们对血液中酒精浓度最大值($Y_{\max}$)及其出现时间(t_{p1})进行求解,利用一次函数极值求解的方法,容易得到

$$t_{p1} = (\ln k - \ln k_0)/(k - k_0)$$

$$Y_{\max} = Y(t_{p1}) = (fnm/V) \cdot (k_0/k)^{-k/(k-k_0)}$$

要使 $Y(t)$ 符合国家标准 Y_0,只需 $Y(t) \leqslant Y_0$,编程搜索即可求解到达安全范围的最少时间 t_1。

②T 时间内等时间间隔 Δt 等量喝酒。考虑到人不可能长时间一直不间断饮酒,这里我们用等时间等量饮酒进行模拟。设 t_{p2} 为等时间间隔喝酒时血液中酒精浓度达最大值的时间。假定喝 r 次,则 $r = T/\Delta t + 1$,每次快速喝

酒酒精量为 ΔD，则 $\Delta D=fnm/r$，建立叠加微分方程模型：

$$\begin{cases} x_i'(t)=-k_0x_i(t) \\ y'_i(t)=k_0x_i(t)-ky_i(t) \\ x_1(0)=\Delta D=W_1,y_1(0)=0=A_1 \\ x_i((i-1)\Delta t)=x_{i-1}((i-1)\Delta t)+\Delta D=W_i \\ y_i((i-1)\Delta t)=y_{i-1}((i-1)\Delta t)=A_i \quad (i\geqslant 2) \end{cases}$$

解得 t 时刻血液中酒精浓度为

$$Y_i(t)=\frac{y_i(t)}{V}=\frac{W_i\cdot k_0}{V\cdot(k_0-k)}\{e^{-k[t-(i-1)\Delta t]}-e^{-k_0[t-(i-1)\Delta t]}\}$$
$$+A_ie^{-k[t-(i-1)\Delta t]}(\mathrm{mg/100\ mL})$$

下面用递推的方法求得 W_i 和 A_i：

$W_1=\Delta D,W_i=W_{i-1}e^{-k_0\Delta t}+\Delta D \quad (2\leqslant i\leqslant r)$；

$A_1=0,A_i=[W_{i-1}k_0/(k_0-k)](e^{-k\Delta t}-e^{-k_0\Delta t})+A_ie^{-k\Delta t} \quad (2\leqslant i\leqslant r)$

由 W_i 和 A_i 的递推式经整理可得它们的显式表达：

$$W_i=\frac{1-e^{-ik_0\Delta t}}{1-e^{-k_0\Delta t}}\Delta D;$$

$$A_i=\frac{\Delta Dk_0}{(k_0-k_1)(1-e^{-k_0\Delta t})}\left[\frac{1-e^{-(i-1)k_0\Delta t}}{1-e^{\ k_0\Delta t}}-\frac{1-e^{-(k_0-k)(i-1)\Delta t}}{1-e^{-(k_0-k)\Delta t}}e^{-(i-1)k_0\Delta t}\right]$$
$$(e^{-k\Delta t}-e^{-k_0\Delta t})$$

式中 $i=1,2,3,\cdots$。

W_i 和 A_i 均已表达出来，那么任意时刻血液中的酒精浓度的表达式可求得。

每次喝酒时，前一次喝的酒都有残余量，则血液中酒精浓度最大值必定是在最后一次饮酒完之后才出现的，由于 $i=r,\Delta t=T/(r-1)$，则 W_i 和 A_i 化为关于 T,r 的函数 A_r',W_r'，那么 t_{p2} 也必定是关于 r 与 T 的函数。利用函数求极值的方法，得

$$t_{p2}(T,r)=\frac{1}{k_0-k}\ln\frac{W_r'k_0^2}{W_r'kk_0+A_r'kk_0-A_r'k_0^2}+T;$$

$$Y'_{\max}=\frac{W_r'k_0}{V}e^{-k_0(t_{p2}-T)}$$

令 $i=r$，代入 $Y_i(t)$ 式得 $Y_i(t)$，只需 $Y_i(t)\leqslant Y_0$。编程搜索即可得在固定 r 和 T 的情况下，饮酒者从开始饮酒算起几小时后能恢复。由

$$\lim_{r\to\infty}W_r=\frac{\Delta D}{1-e^{-k_0\Delta t}};$$

$$\lim_{r\to\infty}A_r=\frac{\Delta Dk_0}{(k_0-k)(1-e^{-k_0\Delta t})(1-e^{-k\Delta t})}(e^{-k\Delta t}-e^{-k_0\Delta t})$$

可以看出，当每次喝酒量一定、间隔时间一定而次数是无穷大时，酒精浓度可以达到稳定。

③天天喝酒模型。设日饮酒量 D，且每天饮酒时刻相等，$y_i(t)$ 为第 i 天 t 时刻血液中酒精含量，$x_i(t)$ 为第 i 天胃肠道中酒精含量，$Y_i(t)$ 为第 i 天胃肠道中酒精浓度，A_i 为第 i 天喝酒前血液中酒精含量，W_i 为第 i 天喝酒前胃肠道中酒精含量。

查阅相关文献知，胃中的酒精 2.5～3 h 之后即被完全吸收，因此 $W_i=D$，所以

$$\begin{cases} x'_i(t)=-k_0x_i(t) \\ y'_i(t)=k_0x_i(t)-ky_i(t) \\ x_i(0)=W_i \\ y_1(0)=0=A_1 \\ y_i(0)=y_{i-1}\Delta t=A_i \quad (i\geqslant 2) \end{cases}$$

解得 t 时刻血液中酒精浓度为

$$Y_i(t)=\frac{y_i(t)}{V}=\frac{D\cdot k_0}{V\cdot(k_0-k)}(\mathrm{e}^{-kt}-\mathrm{e}^{-k_0t})+\frac{A_i}{V}\mathrm{e}^{-kt}(\mathrm{mg/100\ mL})。$$

类似地，

$$A_i=\frac{Dk_0}{(k_0-k)}\frac{1-\mathrm{e}^{-(i-1)k_0\Delta t}}{1-\mathrm{e}^{-k_0\Delta t}}(\mathrm{e}^{-k\Delta t}-\mathrm{e}^{-k_0\Delta t})$$

当 $i\to\infty$ 时，因为 $\exp\{-k_0\Delta t\}$ 非常小，所以

$$\lim_{i\to\infty}A_i=\frac{Dk_0\mathrm{e}^{-k\Delta t}}{(k_0-k)(1-\mathrm{e}^{-k\Delta t})}>A_i$$

用 $\lim\limits_{i\to\infty}A_i$ 替换 A_i，令 $Y_i(t)\leqslant Y_0$，编程搜索即可得天天喝酒的最低恢复时间。

(2)问题的求解

首先，fm/V，k 和 k_0 可用实验数据拟合求得，结果为：

$$k=2.0079,\quad k_0=0.1855,\quad fm/V=51.9304\ \mathrm{mg/100\ mL}$$

由此模型①(一次性快速喝下 n 瓶酒)简化为

$$Y(t)=114.4327\times(\mathrm{e}^{-0.1855t}-\mathrm{e}^{-2.0079t})。$$

①大李超标的原因分析。酒精浓度是否超标主要在于喝酒方式如何。如果是在短时间内喝下去的，可以认为瞬间全部运到消化道中。根据模型②(T 时间内等时间间隔 Δt 等量喝酒)，令喝酒次数为 2 次，时间间隔为 6 h，经计算，这种情况不符合题意。由于 6 点时大李正在接受检查，不可能吃饭；结合模型②，代入数值，解得若大李 7 点吃晚饭，凌晨 2 点检查其血液中酒精浓度为 20.4787 mg/100 mL。如果大李中午和晚上均分 3 次喝，分别间隔 0.1 h 和 0.3 h，血液中酒精浓度的峰值比前一种情况小，但它能长时间维持较高浓

度。结合模型②,可计算两次检查时血液酒精浓度分别为 19.8917 mg/100 mL 和 20.1477 mg/100 mL。

②这里分两种情况进行讨论。第一种情况:在很短的时间内喝下 3 瓶啤酒或低度白酒。

根据模型①,代入数据有:

$$Y(t)=114.4327\times(e^{-0.1855t}-e^{-2.0079t})\leqslant 20,$$

通过编程搜索求解得:$t\geqslant 9.4030$ (h)。

第二种情况:在较长一段时间内喝下 3 瓶啤酒或低度白酒。

我们以 2 h 为例,分别结合模型①,血液中酒精浓度不超过标准需要的最少休息时间(从第一次喝酒算起)与喝酒次数的关系结果如图 6-10 所示。

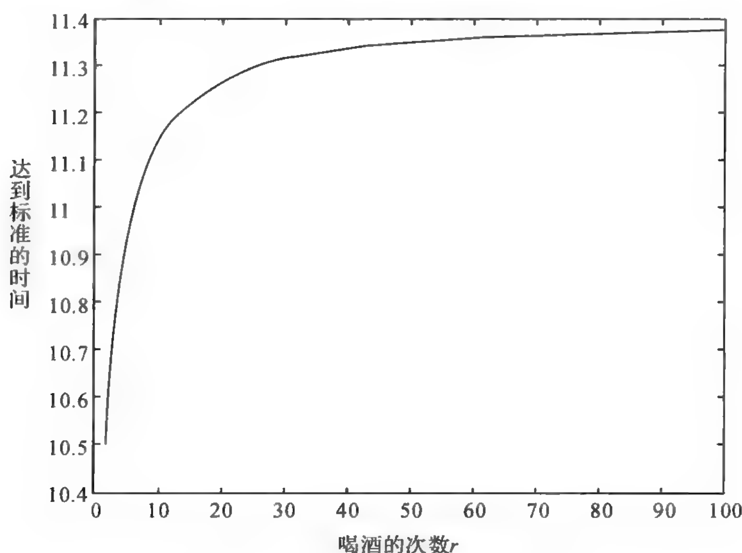


图 6-10 喝酒次数与休息时间

③如果酒是在很短的时间内喝掉。

可根据模型①得出结论:

$$t_{p1}=(\ln k-\ln k_0)/(k-k_0)=1.2217$$

可以看出此种情况下,由于 k_0 和 k 都是常数,所以 t_{p1} 也是定值,它与饮酒量无关。即按这种喝法,血液中酒精浓度总在喝完酒 1.2217 h 后达到最高。

如果酒是在较长的一段时间内喝的,同样运用求导的方法,由模型②求得达到峰值时刻的表达式知,t_{p2} 也与饮酒量无关,仅与 r、T 有关。

④如果每天只有一顿饭喝酒,且时间和饮酒量均固定。

由天天喝酒模型结合编程得喝啤酒的瓶数与酒后需要最少恢复到能开

车状态的时间的关系如表 6-2。

表 6-2　喝的啤酒瓶数与恢复需要的时间 H(小时) 的关系

酒的瓶数	1	2	3	4	5	6	7
H	5.7306	9.4706	11.6606	13.2106	14.4106	15.3906	16.2206

利用表 6-2 我们可以得出如下饮酒策略。由于 H 均小于 24 小时，即第二天的同一时间可以恢复，说明每天可以喝一点酒，但是注意饮酒量不能太大，否则恢复时间过长会影响工作，而且还应注意时间的调整。如果第二天白天需要开车而当天晚上不开车，则晚饭可以喝一点；如果当天晚上要开车，而下午不开车，则可在吃午饭时喝少量的酒。

6.4　传染病问题

6.4.1　概念

(1)概念

传染病(Infectious Diseases)是由各种病原体引起的能在人与人、动物与动物或人与动物之间相互传播的一类疾病。病原体中大部分是微生物，小部分为寄生虫，寄生虫引起的病又称寄生虫病。

传染病的特点是有病原体、有传染性和流行性，感染后常有免疫性。有些传染病还有季节性或地方性。传染病的预防应采取以切断主要传播环节为主导的综合措施。传染病的传播和流行必须具备三个环节，即传染源(能排出病原体的人或动物)、传播途径(病原体传染他人的途径)及易感者(对该种传染病无免疫力者)。若能完全切断其中的一个环节，即可防止该种传染病的发生和流行。各种传染病的薄弱环节各不相同。

(2)特征

①传染特征——流行性、地方性、季节性。

流行性：按传染病传播过程的强度和广度分类。散发是指传染病在人群中散布发生；流行是指某一地区或某一单位在某一时期内某种传染病的发病率，超过了历年同期的发病水平；大流行指某种传染病在一个短时期内迅速传播、蔓延，超过了一般的流行强度；暴发指某一局部地区或单位在短期内突然出现众多的同一种疾病的病人。

地方性：指某些传染病或寄生虫病，其中间宿主受地理条件、气温条件变化的影响，常局限于一定的地理范围内发生。如虫媒传染病、自然疫源性疾病。

季节性:指传染病的发病率在年度内有季节性升高。此与温度、湿度的改变有关。

②传染后特征——免疫性。

传染病痊愈后,人体对同一种传染病病原体产生不感受性,称为免疫。不同的传染病,病后免疫状态有所不同,有的传染病患过一次后可终身免疫,有的还可能会再感染。可分为以下几种感染现象。

再感染:同一传染病在痊愈后,经过一定时间后,被同一种病原体感染。

重复感染:某种疾病在发病中,被同一种病原体再度侵袭而受感染。血吸病、丝虫病、疟疾最为常见。

复发:发病过程已转入恢复期或接近痊愈,而该病原体再度出现并繁殖,原症状再度出现。伤寒最为常见。

再燃:临床症状已缓解,但体温尚未正常而又复上升、症状略见加重者。见于伤寒。

(3)传播

病原体从已感染者排出,经过一定的传播途径,传入易感者而形成新的传染的全部过程。传染病要在某一人群中发生和传播,必须具备传染源、传播途径和易感人群三个基本环节。

①传染源。在体内有病原体生长繁殖并可将病原体排出的人和动物,即患传染病或携带病原体的人和动物。病程的各个时期,病人的传染源作用不同,这主要与病种、排出病原体的数量和病人与周围人群接触的程度及频率有关。多数传染病病人在有临床症状时能排出大量病原体,威胁周围人群,是重要的传染源。如乙型肝炎病人在潜伏期末才具有传染性。

一般说来,病人在恢复期不再是传染源,但某些传染病(伤寒、白喉)的恢复期病人仍可在一定时间内排出病原体,继续起传染源的作用。

患病动物也是人类传染病的传染源。人被患病动物咬伤或接触患病动物的排泄物、分泌物而被感染。人和动物可患同一种病,但病理改变、临床表现和作为传染源的意义不相同。如患狂犬病的狗是该病的传染源之一,但人患此病后不再传染。

②传播途径。指病原体自传染源排出后,在传染给另一易感者之前在外界环境中所行经的途径。一种传染病的传播途径可以是单一的,也可以是多个的。传播途径可分为水平传播和垂直传播两类。

③易感人群。是指人群对某种传染病病原体的易感程度或免疫水平。新生人口增加、易感者的集中或进入疫区、部队的新兵入伍等,易引起传染病流行。

(4)世界主要致死传染病概况

世界卫生组织搜集全球死亡数据,表6-3为2002年世界主要致死疾病,这些疾病都造成至少10万人死亡,1993年数据随表附加以资参考。需特别注意表中前三名致死单一疾病分别是艾滋病、结核和疟疾,尽管几乎所有的疾病所造成的死亡案例皆逐步减少,但死于艾滋者却以四倍成长,而百日咳、小儿麻痹、白喉、麻疹和破伤风等儿童疾病占很大比例,此外儿童也在腹泻、下呼吸道感染中占很大的百分比。

表6-3 2002年和1993年世界主要致死疾病死亡人数

排名	死因	2002年死亡人数(百万)	占所有死亡百分比(%)	1993年死亡人数(百万)	1993年排名
无	所有感染症	14.7	25.9	16.4	—
1	下呼吸道感染	3.9	6.9	4.1	1
2	艾滋病	2.8	4.9	0.7	7
3	肠胃炎	1.8	3.2	3.0	2
4	结核	1.6	2.7	2.7	3
5	疟疾	1.3	2.2	2.0	4
6	麻疹	0.6	1.1	1.1	5
7	百日咳	0.29	0.5	0.36	7
8	破伤风	0.21	0.4	0.15	12
9	脑膜炎	0.17	0.3	0.25	8
10	梅毒	0.16	0.3	0.19	11
11	B型肝炎	0.10	0.2	0.93	6
12～17	六种热带疾病	0.13	0.2	0.53	9、10、16～18

6.4.2 传染病模型

(1)传染病模型1

设某地区共有$n+1$人,最初时刻共有i人得病,t时刻已感染疾病的人数为$i(t)$,假定每一已感染者在单位时间内将疾病传播给k个人(k在医学上被称为该疾病在该地区的传染强度),且设此疾病既不导致死亡也不会康复(注:疾病流传初期的一段较短时间内情况大体如此),则可导出传染病的马尔萨斯模型:

$$\frac{di}{dt}=k\cdot i,\text{初始值 } i(0)=i_0$$

解得：$i(t)=i_0e^{kt}$。

(2)传染病模型 2

记 t 时刻的病人数与易感染人数分别为 $i(t)$ 与 $s(t)$，初始时刻的病人数为 i。根据病人不死也不会康复的假设，可得传染病的逻辑斯蒂克模型：

$$\frac{di}{dt}=k\cdot i\cdot s,\text{其中 } i+s=n+1,\text{初始值 } i(0)=i_0$$

解得：

$$i(t)=\frac{c_0(n+1)e^{k(n+1)t}}{1+c_0e^{k(n+1)t}},\text{其中}:C_0=i_0/(n+1-i_0)$$

统计结果显示，模型 2 预报结果比模型 1 更接近实际情况。医学上称坐标$(i,di/dt)$构成的曲线 $i\sim di/dt$ 为传染病曲线，并称 di/dt 最大值时刻 t_1 为此传染病的流行高峰。令$\frac{d^2i}{dt^2}=0$，得 $t_1=-\frac{\ln c_0}{k(n+1)}$。

(3)传染病模型 3

将人群划分为三类：易感染者、已感染者和已恢复者，分别记 t 时刻的三类人数为 $s(t)$，$i(t)$和 $r(t)$，则可建立下面的三房室模型：

$$\begin{cases}\frac{di}{dt}=ksi-li & (1)\\ \frac{dr}{dt}=li & (2)\\ s(t)+i(t)+r(t)=n+1 & (3)\\ i(0)=i_0,r(0)=0\end{cases}$$

其中 l 称为此传染病在该地区的恢复系数(l 越大恢复越快)。

模型 3 可按如下方式求解。

对(3)式求导，由(1)(2)得：

$$\frac{ds}{dt}=-ksi=-\frac{k}{l}s\frac{dr}{dt}\quad(\text{由}(2)\text{式})$$

解得 $s(t)=s_0e^{-\frac{k}{l}r(t)}$。

记 $\rho=l/k$，则 $s(t)=s_0e^{-\frac{1}{\rho}r(t)}$。

由(1)式可得：

$$\frac{di}{dt}=-\frac{ds}{dt}-li=\frac{ds}{dt}+\frac{\rho ds}{s\,dt}$$

积分得

$$i(t)=i_0+s_0-s(t)=\rho\ln\frac{s(t)}{s_0}$$

从而解得：

$$
\begin{cases}
i(t)=i_0+s_0-s(t)+\rho\ln\dfrac{s(t)}{s_0} \\
s=s_0\mathrm{e}^{-\rho^{-1}r(t)} \\
r(t)=n+1-i(t)-s(t)
\end{cases}
$$

不难验证，当 $t\to+\infty$时，$r(t)$趋向于一个常数，从而解释了医生们发现的现象。

为揭示产生上述现象的原因，将方程组中的第(1)式改写成：

$$\frac{\mathrm{d}i}{\mathrm{d}t}=ki(s-\rho)$$

其中 $\rho=l/k$ 通常是一个与疾病种类有关的较大的常数。容易看出，如果 $s_0\leqslant\rho$，则有 $\mathrm{d}i/\mathrm{d}t<0$，此疾病在该地区流行不起来；如果 $s_0>\rho$，则开始时$\mathrm{d}i/\mathrm{d}t>0$，$i(t)$单调递增。但在 $i(t)$增加的同时，伴随地有 $s(t)$单调递减。当 $s(t)$减少到$\leqslant\rho$时，$\mathrm{d}i/\mathrm{d}t\leqslant0$，$i(t)$开始减小，直至此疾病在该地区消失。鉴于 ρ 在本模型中的作用，医生们称其为此疾病在该区的阀值，ρ 的引入解释了为什么此疾病没有波及该地区的所有人。

综上所述，模型 3 指出了传染病的以下特征。

①当人群中有人得了某种传染病时，此疾病并不一定流传，仅当易受感染的人数与超过阀值时，疾病才会流传起来。

②疾病并非因缺少易感染者而停止传播，相反，是因为缺少传播者才停止传播的(否则将导致所有人得病)。

③种群不可能因为某种传染病而绝灭。

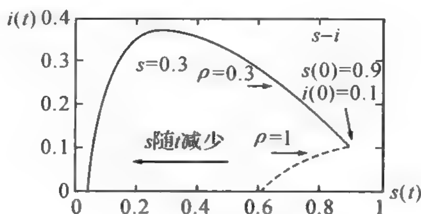


图 6-11　传染病模型曲线

从图 6-11 中可以看出，当 $s>\rho$ 时，$i(t)$是增加的；当 $s<\rho$ 时，$i(t)$是减少的；当 $s=\rho$ 时，$i(t)$达到最大值。

下面对模型进行检验。

医疗机构一般依据 $r(t)$来统计疾病的波及人数，从广义上理解，$r(t)$应为 t 时刻已就医而被隔离的人数，这些人已被隔离，既不再传染给别人，也不再

被别人传染，至于是真的康复了还是死亡了对模型并无影响(注：$i(t)$可视为具有传染性的潜伏期病人)。

注意到 $$\frac{\mathrm{d}r}{\mathrm{d}t}=li=l(n+1-r-s)\text{及 }s=s_0\mathrm{e}^{-\frac{r}{\rho}}$$

可得： $$\frac{\mathrm{d}r}{\mathrm{d}t}=l(n+1-r-s_0\mathrm{e}^{-\frac{r}{\rho}})$$

通常情况下，传染病波及的人数占总人数的百分比不会太大，故一般是小量。利用泰勒公式展开取前三项，最后解得：

$$r(t)=\frac{\rho^2}{s_0}\left[\frac{s_0}{\rho}-1+m\tanh\left(\frac{1}{2}mlt-\varphi\right)\right]$$

其中，$m=\left[\left(\frac{s_0}{\rho}-1\right)^2+\frac{2s_0(n+1-s_0)}{\rho}\right]^{1/2}$，$\varphi=\tanh^{-1}\frac{1}{m}\left[\frac{s_0}{\rho}-1\right]$，$\tanh u=\frac{\mathrm{e}^u-\mathrm{e}^{-u}}{\mathrm{e}^u+\mathrm{e}^{-u}}$为双曲正切函数。

导函数曲线$\frac{\mathrm{d}r}{\mathrm{d}t}=\frac{lm^2\rho^2}{2s_0}\mathrm{sech}^2\left(\frac{1}{2}mlt-\varphi\right)$为一单峰曲线，在医学上被称为疾病传染曲线，可用医疗单位每天实际登录数进行比较拟合得到最优曲线。

思考题

1. 请给广大司机朋友一张饮酒量与驾驶时间表，要求有不同体重、不同酒精浓度、不同饮酒量对应的酒后到能开车的等待时间。要求：先建模，再选择若干情形计算。

2. 甲型 H1N1 流感为急性呼吸道传染病，其病原体是一种新型的甲型 H1N1 流感病毒，在人群中传播。人群对甲型 H1N1 流感病毒普遍易感，并可以人传染人，人感染甲流后的早期症状与普通流感相似，包括发热、咳嗽、喉咙痛、身体疼痛、头痛、发冷和疲劳等，有些还会出现腹泻或呕吐、肌肉痛或疲倦、眼睛发红等。2009 年开始，甲型 H1N1 流感在全球范围内大规模流行。2010 年 8 月，世卫组织宣布甲型 H1N1 流感大流行期已经结束。请收集墨西哥、美国、中国等国家 2009—2010 年那场甲流疑似、确诊、治愈、死亡数据和这些国家所采取的防治措施，绘制它们随时间变化的曲线，对照理论模型，比较这些国家实际数据的异同。

3. 渔场承包问题。人类与自然资源密切相关，许多资源由于掠夺式的过度开发，导致资源日益减少甚至枯竭，这会严重危及人类的生存和发展，渔业作为可再生利用资源，采用何种捕捞方式，才能既保证鱼群的持续稳定发展而不至于枯竭，又要能提供最大数额的产量，是渔业管理中的一个重要问题。

假设我们现在制定的捕捞鳀鱼的策略，设

(1)鱼群按年龄分为四组,称为1龄鱼,…,4龄鱼,平均体重分别为5.07,11.55,17.86,22.99(g),各组自然死亡率均为0.8/年;

(2)捕鱼期只能在每年的前8个月,后4个月为鳀鱼的集中产卵孵化期,平均每条4龄鱼产卵1.109×10^5(个),3龄鱼产卵为其一半,卵孵化后成1龄鱼,成活率为$1.22\times10^{11}/(1.22\times10^{11}+n)$,$n$为产卵总量;

(3)捕鱼采用固定努力量方式,即单位时间捕鱼量与该年龄的鱼总数成正比,比例系数为捕捞强度系数,捕捞只能捕到3龄鱼和4龄鱼,其捕捞强度系数比为0.42∶1。

现在考虑如何实现可持续捕捞,即每年年初各年龄组鱼的数目保持稳定以及在稳定状态下采用何种捕捞强度能使年捕获量最大;渔业公司承包该鱼5年的捕捞业务,开始鱼群分布已知[各年龄组数目分别为:122,29.7,10.1,3.29($\times10^9$)条],该公司应如何决策才能获得最大捕捞总量而不损害鱼群生产能力。

第 7 章

战争与和平——博弈理论

“田忌赛马”的故事家喻户晓，但本章还是从这一故事开始说起。

战国初期，齐国的大将田忌和齐威王约定赛马，商量好先把各自的马分成上、中、下三等。比赛的时候，上马对上马，中马对中马，下马对下马。由于齐威王每个等级的马都比田忌的马强壮，所以比赛了几次，田忌都失败了。后来田忌的好朋友孙膑给他出主意，孙膑先以下等马对齐威王的上等马，自然是田忌输了；第二场比赛，孙膑拿上等马对齐威王的中等马，获胜；第三场比赛，孙膑拿中等马对齐威王的下等马，又战胜了一局。比赛的结果是三局两胜，田忌赢了齐威王。还是同样的马匹，由于调换一下比赛的出场顺序，就转败为胜。

显然，“田忌赛马”不能算作真正的博弈，因为只有一方可以改变策略。假如，后来齐王与田忌又约定赛马，齐王吸取教训，改变策略，结局如何就不一定了。真正的博弈应该是参与各方都不是大傻瓜，都有改变策略的可能，最终胜负要看谁技高一筹。

7.1 什么是博弈

7.1.1 博弈原意

博弈，中文解释是局戏、围棋、赌博。博弈的英文是 gaming，game 或 game playing。Game 的意思是比赛、竞赛、玩耍或赌博等。

(1)局戏、围棋

《论语·阳货》：“饱食终日，无所用心，难矣哉！不有博弈者乎？为之，犹贤乎已。”朱熹集注：“博，局戏；弈，围棋也。”

《汉书·游侠传·陈遵》：“祖父遂，字长子，宣帝微时与有故，相随博弈，数负进。”颜师古注：“博，六博；弈，围棋也。”

(2)赌博

苏轼《策别》十七：“出为盗贼，聚为博弈，群饮于市肆，而叫号于郊野。”

戴名世《财神问对》：“聚为博弈，出为盗贼。”

博弈是中国古代游戏活动的重要组成部分，大体有六博、双陆、打马格、围棋和象棋等项目，是古人展现智慧、运筹争胜的重要方式。

博弈论是两人或多人在平等的对局中各自利用对方的策略变换自己的对抗策略，达到取胜的目的。博弈论思想古已有之，我国古代的《孙子兵法》就不仅是一部军事著作，而且算是最早的一部博弈论著作。博弈论最初主要研究象棋、桥牌、赌博中的胜负问题。

7.1.2 再说田忌赛马

首先假设比赛的时候，齐威王每个等级的马都比田忌的马强壮。上马对上马，中马对中马，下马对下马，齐王的马会全胜。如果不同等级的马比赛，高等级的胜低等级的，如上马胜中马和下马，中马胜下马。现在约定，比赛中赛马出场方案由参与双方自己决定。这里有两种可能细节，一是比赛前出场方案完全预先确定，二是每场比赛后可临时调整下一场的策略。

(1)比赛前出场方案完全确定

假定要求开赛前必须确定所有选手的出场顺序，这是两人博弈，显然可以先将一方的方案固定，譬如田忌的方案固定为{上马、中马、下马}出场顺序。对于齐王来说，共有六种方案可选择：

{上马、中马、下马}，{上马、下马、中马}，{中马、上马、下马}，

{中马、下马、上马}，{下马、上马、中马}，{下马、中马、上马}。

如果齐王的同等级马都优于田忌，那么这六种方案中只有{中马、下马、上马}出场顺序齐王会输。由于齐王预先不知道田忌方案，假如齐王采用抽签方法从六种方案中任取一种，那么抽中{中马、下马、上马}方案的可能性是1/6。也就是说，如果采取完全随机的办法，齐王输的概率是1/6，胜的概率是5/6。

(2)每赛完一场可调整下一场出场方案

假定允许每赛完一场可调整下一场出场方案，过程明显复杂多了。

第一场比赛共有九种可能：

{上马对上马}，{上马对中马}，{上马对下马}，{中马对上马}，{中马对中马}，{中马对下马}，{下马对上马}，{下马对中马}，{下马对下马}。

注意到，齐王只要不出现上马对田忌的下马、中马对田忌的上马、下马对田忌的中马，就赢定了。第一场后，齐王有6/9＝2/3的概率肯定赢了，还有3/9＝1/3的概率未知输赢。如果第一场是齐王上马对田忌下马或中马对田忌上马、下马对田忌中马，则胜负未定。

不妨假设第一场比赛刚好是齐王上马对田忌的下马，虽然齐王先胜一

局，但结局难料。这样，齐王留下中马和下马，田忌是上马和中马。后两场比赛共有两种对阵方式，但其中只有一种有利于齐王，即齐王中马对田忌中马、齐王下马对田忌上马，这样一胜一负，最终齐王胜。而另一种方式是齐王下马对田忌中马、齐王中马对田忌上马，齐王连输两局，最终负。因此，最后齐王与田忌胜负各一半。

同样地，第一场比赛刚好是齐王中马对田忌的上马或齐王下马对田忌中马，齐王先负一局。后两局，齐王连胜两局和一胜一负的可能性还是各一半，最后胜负也是各一半。

总体上，齐王胜的概率是：$2/3+1/3\times1/2=5/6$。与前面小节结论一致。

事实上，在完全公平的博弈中，参与双方胜负的概率是一定的，不管中间的程序过程是什么。在两军对垒的战场上，如果双方指挥员同样智慧，那么士兵多、装备好的一方胜的概率更大。但历史上总有一些以少胜多、以弱胜强的战例。如官渡之战，曹操以两万左右的兵力，出奇制胜，击破袁军十万。此战曹操善择良策，攻守相济，屡出奇兵，巧施火攻，焚烧袁军粮草，对获胜起重大作用，集中体现曹操卓越的用兵谋略和指挥才能，是中国历史上以少胜多的著名战例。反观袁绍，内部不和，又骄傲轻敌，刚愎自用，屡拒部属的正确建议，迟疑不决，一再地丧失良机。终致粮草被烧，后路被抄，军心动摇，内部分裂，而全军溃败。

7.2 现代博弈理论

7.2.1 现代博弈论

现代数学中有博弈论，亦名“对策论”“赛局理论”，属应用数学的一个分支，表示在多决策主体之间行为具有相互作用时，各主体根据自己掌握的信息及对自身能力的认知，做出有利于自己的决策的一种行为理论。目前在生物学、经济学、国际关系、计算机科学、政治学、军事战略和其他很多学科都有广泛的应用。博弈论主要研究公式化了的激励结构间的相互作用，是研究具有斗争或竞争性质现象的数学理论和方法，也是运筹学中一个重要分支。

早在20世纪初，塞梅鲁(Zermelo)、鲍罗(Borel)和冯·诺依曼(Von Neumann)已经开始研究博弈的准确的数学表达，直到1939年，冯·诺依曼遇到经济学家奥斯卡·摩根斯特恩(Oskar Morgenstern)并与其合作才使博弈论进入经济学的广阔领域。1944年，冯·诺依曼与奥斯卡·摩根斯特恩合著的巨作《博弈论与经济行为》出版，标志着现代系统博弈理论的初步形成。

该书提出的标准型、扩展型和合作型博弈模型解的概念和分析方法，奠定了这门学科的理论基础。1950年和1951年纳什关于非合作博弈论的两篇重要论文，彻底改变了人们对竞争和市场的看法。他证明了非合作博弈及其均衡解，并证明了均衡解的存在性，即著名的纳什均衡，从而揭示了博弈均衡与经济均衡的内在联系。纳什的研究奠定了现代非合作博弈论的基石，后来的博弈论研究基本上都是沿着这条主线展开的。

7.2.2 博弈分类

(1)合作博弈和非合作博弈

根据是否可以达成具有约束力的协议，博弈分为合作博弈和非合作博弈。

①合作博弈亦称正和博弈，是指博弈双方的利益都有所增加，或者至少是一方的利益增加，而另一方的利益不受损害，因而整个社会的利益有所增加。合作博弈研究人们达成合作时如何分配合作得到的收益，即收益分配问题。合作博弈采取的是一种合作的方式，或者说是一种妥协。合作博弈之所以能够增进妥协双方的利益以及整个社会的利益，就是因为合作博弈能够产生一种合作剩余。这种剩余就是从这种关系和方式中产生出来的，且以此为限。至于合作剩余在博弈各方之间如何分配，取决于博弈各方的力量对比和技巧运用。因此，妥协必须经过博弈各方的讨价还价，达成共识，进行合作。在这里，合作剩余的分配既是妥协的结果，又是达成妥协的条件。

合作博弈强调的团体理性(Collective Rationality)，是效率、公平、公正。

合作博弈存在的两个基本条件：一是对联盟来说，整体收益大于其每个成员单独经营时的收益之和；二是对联盟内部而言，应存在具有帕累托改进性质的分配规则，即每个成员都能获得比不加入联盟时多一些的收益。

合作按照合作之后的收益变化可分为本质性的合作和非本质性的合作。如果合作后收益有所增加，则此合作博弈是本质性的，即存在有净增收益的联盟；如果合作后收益没有增加甚至下降，则为非本质性合作。例如，我国现存的一些低效率、名不副实的集团及一些经济合作组织可看做是非本质性合作，因为这类合作组织并没有真正发挥合作优势，并没有创造出比不合作时更大的社会经济效益。

②非合作博弈是指一种参与者不可能达成具有约束力的协议的博弈类型，这是一种具有互不相容味道的情形。非合作博弈研究人们在利益相互影响的局势中如何选择决策使自己的收益最大，即策略选择问题。

负和博弈与零和博弈统称为非合作博弈，正和博弈亦称为合作博弈。

纳什均衡(Nash Equilibrium)，又称为非合作博弈均衡，是博弈论的一个

重要术语，以约翰·纳什命名。1948年，约翰·纳什作为年轻数学博士生进入普林斯顿大学。其研究成果见于题为《非合作博弈》(1950)的博士论文。该博士论文引出了《n人博弈中的均衡点》(1950)和《非合作博弈》(1951)两篇论文的发表。纳什在上述论文中，介绍了合作博弈与非合作博弈的区别。他对非合作博弈的最重要贡献是阐明了包含任意人数局中人和任意偏好的一种通用解概念，也就是不限于两人零和博弈。该解概念后来被称为纳什均衡。

人们分工与交换的经济活动就是合作性的博弈，而囚徒困境以及公共资源悲剧都是非合作性的博弈。

(2)完全信息博弈与不完全信息博弈

①完全信息博弈：每一参与者都拥有所有其他参与者的特征、策略集及得益函数等方面的准确信息的博弈。

关于完全信息博弈的最早结果出现在20世纪50年代，该定理认为，重复博弈的策略均衡结局与一次性博弈中的可行的个体理性结局恰好相一致，这个结局可被视为把多阶段非合作行为与一次性博弈的合作行为联系在一起。或者可以说，只要行为人有足够的耐心，任何满足个体理性的可行支付都可以通过一个特定的子博弈精炼均衡达到。然而，虽然所有可行的个体理性结局确实代表了合作博弈的解观点，但是它不能够提供相关信息，并且是相当模糊的。

完全信息博弈又可细分为完全信息动态博弈和静态博弈。

完全信息动态博弈，是指博弈中信息是完全的，即参与双方都对他参与的战略空间和战略组合下的支付函数有完全的了解，但行动是有先后顺序的，后行动者可以观察到先行动者的行动，了解先行动者行动的所有信息，而且一般都会持续一个较长时期。

例如"以牙还牙"策略。在定价博弈中，"以牙还牙"策略是指：一家厂商定高价，只要对方继续合作也定高价，那么这家厂商就会一直保持高价；一旦对方定低价，那么该厂商也会定低价；如果对方以后决定合作并再提高价格，该厂商也会提高价格。

又例如"序列博弈"，是指参与者选择策略有时间先后的博弈形式。它是一种较为典型的动态博弈，一方在决策时，会考虑到另一方的反应行为，并在这种考虑基础上进行自己的当前决策。如，在得知厂商一的最佳策略是选择生产少糖型可乐，厂商二则生产多糖型可乐。

完全信息静态博弈中各博弈方同时决策，且所有博弈方对博弈中的各种情况下的策略及得益都完全了解。

例如"上策均衡"，是指对某博弈方来说，不管其他博弈方采取什么策略，

他所采取的能给他带来最大得益的策略。如厂商 A 和 B 的上策都是做广告，那么上策均衡也是两家厂商都选择做广告的策略。

还有“纳什均衡”，指的是在给定竞争对手的选择行为后，博弈方选择了它所能选择的最好的策略(或采取了它所能采取的最好的行动)。

②不完全信息博弈：参与人并不完全清楚有关博弈的一些信息。如大多数纸牌游戏是不完全信息博弈。当然，如果对博弈对手一无所知，那么，也就无从博弈。现实生活中，大多数情况下，虽然对于对手的一些特征不完全了解，但总不至于一无所知。例如，打牌时，虽然不知道对手具体拿什么牌，但根据自己的牌，还是可以估计对手的牌的，而且，随着牌局的展开，人们会不断改进这些估计，这些估计可以用数学上的“概率分布”来表示。

信息不完全静态博弈，是指至少某一个局中人不完全了解另一个局中人的特征，即不知道某一参与人的真实类型，但是知道每一种类型出现的概率。

不完全信息静态博弈则是以贝叶斯均衡等理论完成对混合策略的重新解释，不完全信息动态博弈则是以完美贝叶斯均衡为核心概念的信号博弈。

(3)静态博弈与动态博弈

①静态博弈是指在博弈中，参与人同时选择或不同时选择，但后行动者并不知道先行动者采取什么样的具体行动。对双方来说，都容易形成混沌的行为重组，由于规则的严密与精细，任何人因时间、资金、心理等问题，致使在多次均衡后直到不明不白地大输。参与静态博弈和动态博弈的大部分都是这种人。

②动态博弈是指在博弈中，两个参与人有行动的先后顺序，且后行动者能够观察到先行动者所选择的行动。

7.2.3 博弈论与诺贝尔奖

从 1994 年诺贝尔经济学奖授予 3 位博弈论专家开始，共有 5 届的诺贝尔经济学奖与博弈论的研究有关，分别为：1994 年、1996 年、2001 年、2005 年、2007 年。作为一门工具学科能够在经济学中如此广泛运用并得到学界垂青实为罕见。

7.3 博弈论的基本概念

(1)决策人

在博弈中率先作出决策的一方，这一方往往依据自身的感受、经验和表面状态优先采取一种有方向性的行动。

(2)对抗者

在博弈两人对局中行动滞后的那个人,与决策人要作出基本反面的决定,并且他的动作是滞后的、默认的、被动的,但最终占优。他的策略可能依赖于决策人劣势的策略选择,占去空间特性,因此对抗是唯一占优的方式,实为领导人的阶段性终结行为。

(3)局中人

在一场竞赛或博弈中,每一个有决策权的参与者成为一个局中人。只有两个局中人的博弈现象称为“两人博弈”,而多于两个局中人的博弈现象称为“多人博弈”。

(4)策略

一局博弈中,每个局中人都有选择实际可行的完整的行动方案,即方案不是某阶段的行动方案,而是指导整个行动的一个方案,一个局中人的一个可行的全局筹划的一个行动方案,称为这个局中人的一个策略。如果在一个博弈中局中人都总共有有限个策略,则称为“有限博弈”,否则称为“无限博弈”。

(5)得失

一局博弈结局时的结果称为得失。每个局中人在一局博弈结束时的得失,不仅与该局中人自身所选择的策略有关,而且与全局中人所取定的一组策略有关。所以,一局博弈结束时每个局中人的“得失”是全体局中人所取定的一组策略的函数,通常称为支付(payoff)函数。

(6)次序

各博弈方的决策有先后之分,且一个博弈方要作不止一次的决策选择,就出现了次序问题;其他要素相同、次序不同,博弈就不同。

(7)均衡

均衡是平衡的意思,在经济学中,均衡即相关量处于稳定值。在供求关系中,某一商品市场如果在某一价格下,想以此价格买此商品的人均能买到,而想卖的人均能卖出,此时我们就说,该商品的供求达到了均衡。

纳什均衡(Nash Equilibrium):在一策略组合中,所有的参与者面临这样一种情况,当其他人不改变策略时,他此时的策略是最好的。也就是说,此时如果他改变策略他的支付将会降低。在纳什均衡点上,每一个理性的参与者都不会有单独改变策略的冲动。纳什均衡点存在性证明的前提是“博弈均衡偶”概念的提出。所谓“均衡偶”是在二人零和博弈中,当局中人 A 采取其最优策略 a^*,局中人 B 也采取其最优策略 b^*;如果局中人 B 仍采取 b^*,而局中人 A 却采取另一种策略 a,那么局中人 A 的支付不会超过他采取原来的策略 a^* 的支付。这一结果对局中人 B 亦是如此。

这样,"均衡偶"的明确定义为:一对策略 a^*(属于策略集 A)和策略 b^*(属于策略集 B)称之为均衡偶,对任一策略 a(属于策略集 A)和策略 b(属于策略集 B),总有:偶对$(a,b^*)\leqslant$ 偶对$(a^*,b^*)\geqslant$ 偶对(a^*,b)。

对于非零和博弈也有如下定义:一对策略 a^*(属于策略集 A)和策略 b^*(属于策略集 B)称为非零和博弈的均衡偶,对任一策略 a(属于策略集 A)和策略 b(属于策略集 B),总有:对局中人 A 的偶对$(a,b^*)\leqslant$ 偶对(a^*,b^*);对局中人 B 的偶对$(a^*,b)\leqslant$ 偶对(a^*,b^*)。

有了上述定义,就立即得到纳什定理:

任何具有有限纯策略的两人博弈至少有一个均衡偶。这一均衡偶就称为纳什均衡点。

纳什均衡点概念提供了一种非常重要的分析手段,使博弈论研究可以在一个博弈结构里寻找比较有意义的结果。

但纳什均衡点定义只局限于任何局中人不想单方面变换策略,而忽视了其他局中人改变策略的可能性,因此,在很多情况下,纳什均衡点的结论缺乏说服力,研究者们形象地称之为"天真可爱的纳什均衡点"。塞尔顿(R. Selten)在多个均衡中剔除一些按照一定规则不合理的均衡点,从而形成了两个均衡的精炼概念:子博弈完全均衡和颤抖的手完美均衡。

7.4 博弈案例

7.4.1 囚徒困境

在博弈论中,含有占优战略均衡的一个著名例子是由塔克给出的"囚徒困境"(Prisoner's Dilemma)博弈模型,现在几乎所有的博弈论教科书上都会讲"囚犯的两难处境"的例子,每本书上的例子都大同小异。本文介绍下面一个相对动听的故事,摘自孙健发表在《北京晚报》的文章——《〈美丽心灵〉原型小约翰·纳什的一生》。

话说有一天,一位富翁在家中被杀,财物被盗。警方在此案的侦破过程中,抓到两个犯罪嫌疑人——斯卡尔菲丝和那库尔斯,并从他们的住处搜出被害人家中丢失的财物。但是,他们矢口否认曾杀过人,辩称是先发现富翁被杀,然后只是顺手牵羊偷了点儿东西。于是警方将两人隔离,分别关在不同的房间进行审讯。由地方检察官分别和每个人单独谈话。检察官说,"由于你们的偷盗罪已有确凿的证据,所以可以判你们 1 年刑期。但是,我可以和你做个交易。如果你单独坦白杀人的罪行,我只判你 3 个月的监禁,但你的同伙要被判 10 年刑。如果你拒不坦白,而被同伙检举,那么你就将被判 10 年

刑，他只判 3 个月的监禁。但是，如果你们两人都坦白交代，那么，你们都要被判 5 年刑。”斯卡尔菲丝和那库尔斯该怎么办呢？

表 7-1 给出了这个博弈的支付矩阵。

表 7-1 囚徒困境博弈

B \ A	坦白	抵赖
坦白	−5，−5	0.25，−10
抵赖	−10，0.25	−1，−1

他们面临着两难的选择——坦白或抵赖。显然最好的策略是双方都抵赖，结果是大家都只被判 1 年。但是由于两人处于隔离的情况下无法串供。所以，按照亚当·斯密的理论，每一个人都是从利己的目的出发，他们选择坦白交代是最佳策略。因为坦白交代可以期望得到很短的监禁 ——3 个月，但前提是同伙抵赖，显然要比自己抵赖去坐 10 年牢好。这种策略是损人利己的策略。不仅如此，坦白还有更多的好处。如果对方坦白了而自己抵赖了，那自己就得坐 10 年牢。太不划算了！因此，在这种情况下还是应该选择坦白交代，即使两人同时坦白，至多也只判 5 年，总比被判 10 年好吧。所以，两人合理的选择是坦白，原本对双方都有利的策略（抵赖）和结局（被判 1 年刑）就不会出现。这样两人都选择坦白的策略以及因此被判 5 年的结局被称为“纳什均衡”，也叫非合作均衡。因为，每一方在选择策略时都没有“共谋”（串供），他们只是选择对自己最有利的策略，而不考虑社会福利或任何其他对手的利益。也就是说，这种策略组合由所有局中人（也称当事人、参与者）的最佳策略组合构成。没有人会主动改变自己的策略以便使自己获得更大利益。

“囚徒的两难选择”有着广泛而深刻的意义。个人理性与集体理性的冲突，各人追求利己行为而导致的最终结局是一个“纳什均衡”，也是对所有人都不利的结局。他们两人都是在坦白与抵赖策略上首先想到自己，这样他们必然要服长的刑期。只有当他们都首先替对方着想时，或者相互合谋（串供）时，才可以得到最短时间的监禁的结果。“纳什均衡”首先对亚当·斯密的“看不见的手”的原理提出挑战。按照斯密的理论，在市场经济中，每一个人都从利己的目的出发，而最终全社会达到利他的效果。这位经济学专家在《国富论》中的名言：“通过追求（个人的）自身利益，他常常会比其实际上想做的那样更有效地促进社会利益。”从“纳什均衡”引出了“看不见的手”的原理的一个悖论：从利己目的出发，结果损人不利己，既不利己也不利他。两个囚徒的命运就是如此。从这个意义上说，“纳什均衡”提出的悖论实际上动摇了西方经济学的基石。因此，从“纳什均衡”中我们还可以悟出一条真理：合作

是有利的“利己策略”。但它必须符合以下黄金律:按照你愿意别人以对你的方式来对别人,但只有他们也按同样方式行事才行。也就是中国人说的“己所不欲勿施于人”。但前提是人所不欲勿施于我。其次,“纳什均衡”是一种非合作博弈均衡,在现实中非合作的情况要比合作的情况普遍。所以“纳什均衡”是对冯·诺依曼和摩根斯特恩的合作博弈理论的重大发展,甚至可以说是一场革命。

从“纳什均衡”的普遍意义中我们可以深刻领悟司空见惯的经济、社会、政治、国防、管理和日常生活中的博弈现象。我们可列举出许多类似于“囚徒的两难处境”这样的例子。如价格战、军备竞赛、污染等。一般的博弈问题由三个要素构成:局中人(players)又称当事人、参与者、策略等的集合,策略(strategies)集合以及每一对局中人所做的选择和赢得(payoffs)集合。其中所谓赢得是指如果一个特定的策略关系被选择,每一局中人所得到的效用。所有的博弈问题都会遇到这三个要素。

7.4.2 情侣博弈

(1)麦琪的礼物

美国短篇小说家欧·亨利为我们描述了一个普通却内涵深刻的故事——《麦琪的礼物》。男主人公吉姆是一位薪金仅够维持生活的小职员,女主人公德拉是一位贤惠善良的主妇。他们的生活贫穷,但吉姆和德拉各自拥有一样极珍贵的宝物——吉姆有祖传的一块金表,就算“地下室堆满金银财宝、所罗门王又是守门人的话,每当吉姆路过那儿,准会摸出金表,好让那所罗门王忌妒得吹胡子瞪眼睛”;德拉一头美丽的瀑布般的秀发,可以“使那巴示女王的珍珠宝贝黯然失色”。为了能在圣诞节送给对方一件礼物,吉姆卖掉了他的金表为德拉买了一套“纯玳瑁做的,边上镶着珠宝”的梳子;德拉卖掉了自己的长发为吉姆买了一条白金表链。他们都为对方舍弃了自己最宝贵的东西,而换来的礼物却因此变得毫无价值了。

虽然双方的出发点都是送给对方效用最大的东西,可是最终的结果却事与愿违。出现这个结果的原因是什么呢?怎么样的策略选择才能使双方都受益呢?这是在读完故事之后值得思考的问题。

(2)完全信息静态博弈视角下的情侣博弈

在这个故事中,我们不妨设想,如果吉姆卖了金表,德拉卖了秀发,双方的效用都为-1;吉姆没有卖掉金表,德拉没有卖掉秀发,双方效用都为2;吉姆卖了金表,德拉没有卖掉秀发,各自效用分别为0和1;吉姆没有卖金表,德拉卖掉了秀发,各自效用为1和0。

表 7-2 麦琪的礼物博弈矩阵

吉姆＼德拉	卖	不卖
卖	−1,−1	0,1
不卖	1,0	2,2

从这个博弈矩阵(表 7-2)中我们可以看出,当吉姆选择卖时,德拉的最优策略是不卖;当吉姆选择不卖时,德拉的最优策略仍然是不卖,也就是说无论吉姆怎么选择,德拉选择不卖都将是最优的。同理,无论德拉怎么选择,对于丈夫吉姆来说,不卖始终是他的最优策略。所以,对于这个博弈来说,(不卖,不卖)是最终的纳什均衡解,也就是无论对方采取什么策略我总是采取不卖这个策略是最好的,而且是严格好的。

根据经济学的假设,经济人都是很理性的,上面这个博弈的结果应该选择(不卖,不卖),因为理性人总是从自己的利益出发,居民想生活地好一些,企业想赚钱,可以说都是天经地义的事情,那么对于这个博弈来说,从自身利益的最大化考虑,选择(不卖,不卖)会是最优的结果,双方都可得到最大的效用。那为什么在现实生活中还会出现类似于吉姆和德拉这对夫妻这样的非理性行为呢?简单地说,是因为我们得到的纳什均衡解是不稳定的。

首先,博弈论和经济学其他的分支不一样,它常常不是指出“一定会”怎样,而是指出“可能会”怎样,从“一定会”怎样变成“可能会”怎样,形式上退了一步,但是更有价值,博弈论讲的是真话,不是大话,真话总是具有参考价值的。

其次情侣博弈的实际结局,受别的因素影响较大,可以说是由别的因素来决定的,比方说在上面的博弈中,如果那天是德拉的生日,也许吉姆就会卖了金表,而德拉没卖,出现(卖,不卖)的结果,同样在其他的环境中还会出现(不卖,卖),(卖,卖)的结果。这样的结果,可以说也是合情合理的。

最后,现实中的博弈还受到社会舆论的影响,如果社会舆论更偏重感性,偏重悲剧美,无私奉献,那么在博弈中理性人的假设就会不成立,人们博弈时会更多地从感性、从舆论宣传的道德角度来采取行动。

所以,很多因素会影响博弈的最终结果,圣诞节的礼物只是其中的一个典型罢了,上面的三种情况,不需要全部出现,只要有一种出现,最终的结果都有可能会改变,博弈的均衡解由于受很多非预期因素的影响,会出现不稳定的情况。

那么从博弈论的角度,从理性人的角度思考这个问题,如何才能避免出现对双方都不利的情况呢?情侣博弈从小处说,影响了夫妻双方的利益,往

大处说，可以用来描述友好企业或者有互补性关系的企业之间的合作。可以试想一下，如果双方能够事先沟通一下，那么情况会不会有所改变呢？也许会出现先动优势，即双方都得到好处，但是先决策的一方可能得益多一些。

(3)完全信息动态博弈视角下的情侣博弈

如果德拉先行动，并且告诉吉姆已经给他买了表链。或许吉姆早就想卖了金表来买梳子，但在这种情况下，他就不会再把金表卖了去买梳子。对于德拉来说也是一样，如果吉姆先行动，并且告诉德拉，德拉还会卖掉秀发吗？在这样的情况下，我们假设吉姆先行动，他买了梳子，此时他有两种选择，告诉德拉和不告诉德拉，而德拉也有两种选择，卖和不卖。用博弈树来分析这个现象(见图7-1)。

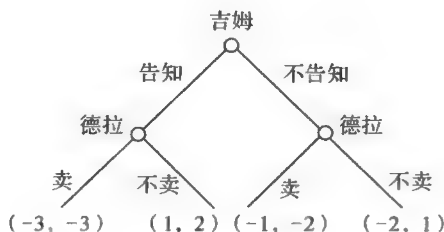


图7-1 麦琪的礼物动态博弈

根据逆向归纳法，不难得出，这个博弈的均衡解是(告知，不卖)，此时的收益是(1,2)。这样的策略组合对双方都是有利的。但是如果这对夫妻都是非常较真的人，喜欢强行承诺，比如吉姆是大男子主义者，德拉是女权主义者，双方都很强势，双方都坚持要为对方买东西。这样的结果将会是(卖，卖)这个劣策略，也就是说无法达到帕累托最优的纳什均衡局面。但如果双方都只是在口头上严词相对，内心反而是相互体谅对方的，最后进行策略选择的时候都作出让步，结果又不一样了。

所以问题的关键是一方在谈判(男女双方协商时本质上就是一种讨价还价的谈判)中做出承诺对于对方的可信度，也就是承诺方是否能够证明：除了遵守承诺以外，其他的选择并不是更痛苦。比如说，德拉能够向吉姆证明：即使她单方面卖掉秀发，不管吉姆如何决定，也能获得极大的满足感。这个时候，德拉所作的一定要卖头发的承诺就是可信的，那么吉姆就不会去卖掉金表了。

7.4.3 基于斗鸡博弈的红绿灯时间设置

(1)斗鸡博弈

考察一个完全信息条件下双人静态博弈，其中参与人A拥有2个纯战

略，记为 $i=1,2$，参与人 B 亦拥有 2 个纯战略，记为 $j=1,2$，支付矩阵如下：

$$U=\begin{bmatrix} a_{11},b_{11} & a_{12},b_{12} \\ a_{21},b_{21} & a_{22},b_{22} \end{bmatrix}$$

式中，a_{ij} 与 b_{ij} 分别是战略组合 (i,j) 所对应的参与人 A 与 B 的支付，$i,j=1,2$。假设支付矩阵满足如下约束条件：

$$a_{21}>a_{11},\quad a_{12}>a_{22},\quad b_{12}>b_{11},\quad b_{21}>b_{22}$$

则可称该博弈为斗鸡博弈。斗鸡博弈存在(1,2)与(2,1)这两个纯战略纳什均衡，分别对应两种不同的帕累托有效状态。斗鸡博弈还存在一个混合战略纳什均衡$((p^*,1-p^*),(q^*,1-q^*))$，这里

$$p^*=(b_{22}-b_{21})/(b_{11}+b_{22}-b_{12}-b_{21})$$
$$q^*=(a_{22}-a_{12})/(a_{11}+a_{22}-a_{12}-a_{21})$$

表示参与人 A 以 p^* 的概率选择第 1 种纯战略，以 $1-p^*$ 的概率选择第 2 种纯战略，参与人 B 则以 q^* 的概率选择第 1 种纯战略，以 $1-q^*$ 的概率选择第 2 种纯战略。

斗鸡博弈的均衡并不唯一，因此无法保证可预期的博弈结果，而且，斗鸡博弈的两个参与人在纯战略均衡状态下的利益具有对立性，一方得利则一方受损，因此，斗鸡博弈描述了一种狭路相逢勇者胜的对局。由于其特殊性，斗鸡博弈常用以解释社会学、财政学中诸多现实问题。

(2)争道问题的斗鸡博弈表示

斗鸡博弈在现实生活中不乏例证，争道问题即为其典型表现。考虑行进方向相冲突的车辆或行人，比如东西向与南北向的车辆或行人，这两类博弈参与人都以前进或等待作为可供选择的战略，相应的支付函数如表 7-3 所示。联系到现实中可能发生的一般性结果，博弈参与人的支付组合满足上小节的约束条件，或者说，这就是一个斗鸡博弈。

表 7-3　争道斗鸡博弈

东西 \ 南北	前进	等待
前进	a_{11}，b_{11}	a_{12}，b_{12}
等待	a_{21}，b_{21}	a_{22}，b_{22}

争道斗鸡博弈存在两个不同的纯战略纳什均衡：东西向参与人前进而南北向参与人等待或东西向参与人等待而南北向参与人前进。该博弈还存在一个混合战略纳什均衡，双方以一定概率随机选择两种纯战略。

尽管斗鸡博弈已从理论上描述了争道问题，我们还需要联系实际情况对博弈支付函数进行具体赋值。就一般情况而言，选择等待则需要支付时间成

本。若撇掉其他因素，则支付可表示为等待时间的一个函数。另外，在双方都选择前进情况下，支付项 a_{11} 与 b_{11} 的取值将联系于可能发生的交通事故风险。

由于均衡的多重性，博弈机制不一定能实现可预期的结果。为保证有效的博弈均衡，争道博弈需要第三方予以规范，作为城市交通文明的标志，红绿灯可行使这一职能。红绿灯在规范行人与车辆行为方面确有其功效，但其约束力究竟是来自参与人的自觉性还是公权力，值得进一步探讨。本项研究拟从经济人的角度考察红绿灯时间比重设置问题。

(3)斗鸡博弈与交通灯时间设置

由于争道问题具有斗鸡博弈性质，不能保证可预期的结果，更不能保证双方皆满意的博弈状态，因此，需要有第三方参与协调，红绿灯得以设置。红绿灯设置是一个系统性问题，涉及红绿灯设置地点、红绿灯空间间隔、红灯与绿灯时间间隔等诸多细节，在这里我们仅关注红绿灯交替时间或绿信比的合理性设置。

如前所述，若采用斗鸡博弈描述争道问题，则红绿灯设置的目的是确保纯战略纳什均衡的出现，红绿灯的变化则表征斗鸡博弈两个纯战略纳什均衡的变化，进一步地，如果斗鸡博弈参与人的支付满足概率约束条件 $p^*=1-q^*$，则斗鸡博弈的混合战略纳什均衡可对应一个符合参与人理性的红绿灯时间间隔。

上述观点可以如此解释：考虑车辆或行人非预谋地出现在这一对局，并未曾预料到可能的红绿灯状态，对于他们而言，红绿灯状态是一个数学意义上的随机事件，而红灯、绿灯各自所占时间比重则给出了该随机事件的概率分布，这相当于给车辆或行人定义了一个混合战略。因此，我们可从红绿灯时间设置与斗鸡博弈混合战略均衡之间的一致性考察红绿灯时间比重设置的合理性。

与表 7-3 所定义的争道博弈相区别，处于交通监管下的争道博弈如表 7-4 所示，红绿灯状态定义了参与人的等待或前进两种不同行动，而支付函数因受到监管的影响也有所不同，体现出行政干扰的特点。

表 7-4　监管下的斗鸡博弈

东西 \ 南北	绿灯	红灯
绿灯	a_{11}，b_{11}	a_{12}，b_{12}
红灯	a_{21}，b_{21}	a_{22}，b_{22}

红绿灯设置应该从经济人基本假设出发，具有博弈均衡特性，特别地，红绿灯时间设置应与斗鸡博弈混合战略均衡保持一致性。单位时间内的红绿灯时间间隔可以采用比例方式，该种时间比例以强制方式定义了博弈参与人

的一个混合战略组合$((p,1-p),(1-q,q))$。进一步，理想的红绿灯时间设置应当确保斗鸡博弈混合战略纳什均衡$((p^*,1-p^*),(q^*,1-q^*))$的出现。综上，我们要求$p^*=1-q^*$，即

$$p^*=(b_{22}-b_{21})/(b_{11}+b_{22}-b_{12}-b_{21})=(a_{11}-a_{21})/(a_{11}+a_{22}-a_{12}-a_{21})$$

或者

$$p^*/(1-p^*)=(b_{22}-b_{21})/(b_{11}-b_{12})=(a_{11}-a_{21})/(a_{22}-a_{12})$$

显然，如果要用斗鸡博弈解释交通灯时间间隔设置问题，必须选择其中某些支付变量作为交通监管媒介，承载交通监管行为的直接影响。通过对支付表的考察，交通监管行为至少可施诸如下两个变量：

①a_{11}。若东西方向车辆或行人闯红灯，应该惩戒其违规行为，这将影响a_{11}的取值。

②b_{11}。若南北方向车辆或行人闯红灯，应该惩戒其违规行为，这将影响b_{11}的取值。

若监管行为只影响a_{11}与b_{11}的取值，则可根据刚才式子得到a_{11}与b_{11}之间的函数关系，该种函数关系是监管合理性的基本前提。为简单起见，令$a_{21}=b_{12}=-1$，则前式简化为

$$(b_{22}-b_{21})/(b_{11}+1)=(a_{11}+1)/(a_{22}-a_{12})$$

根据斗鸡博弈假设当有$(b_{22}-b_{21})(a_{22}-a_{12})>0$，因此，$a_{11}$与$b_{11}$的调整方向应该是相反的，如果$a_{11}$减少则$b_{11}$当增加，反之亦然。

(4)规范车辆争道行为的红绿灯设置

如前所述，要使得斗鸡博弈混合战略纳什均衡与红绿灯时间比重相互协调，博弈参与人的支付函数必须满足给定条件。若不考虑实际背景，满足该条件的支付函数肯定是存在的。若采用斗鸡博弈设置交通灯时间，则需要重新考量支付函数的现实存在性。

交通灯时间设置体现的是一种利益关系，涉及经济、社会、文化等诸多因素，这些因素未必都可量化，或者说，要实现交通灯时间设置的完全合理性几乎不可能，但这不应该成为随意设置城市交通灯时间的理由。我们的基本观点是，交通灯设置应该反映车辆或行人之间的经济利益关系，以此为出发点，可重新设计斗鸡博弈支付函数，进而利用前述方法确定合理的交通灯时间设置。

考虑一个十字路口的红绿灯时间设置问题，我们认为这种地点的红绿灯主要用以规范不同行进方向的车辆行为，我们给出如下假设：

假设1　在无交通监管、交通事故前提下，车辆支付与时间成比例关系。

假设2　给定单位时间内东西方向车流量与南北方向车流量之比为

常数。

假设 3　交通监管较为宽松，容许车辆绿灯状态下的滞留。

在假设 3 下，a_{22} 与 b_{22} 取值不受交通监管行为影响。根据这 3 个假设，我们可对斗鸡博弈支付矩阵赋值。首先，车辆等待成本依赖于时间长短，因此，若面向东西方向车辆的红、绿灯时间间隔比为$(1-p,p)$，面向南北方向车辆的红、绿灯时间间隔比为$(p,1-p)$，则有

$$a_{22}/a_{21}=b_{12}/b_{22}=p/(1-p),\quad b_{22}/b_{21}=a_{12}/a_{22}=p/(1-p)$$

其次，设单位时间内南北方向车流量与东西方向车流量之比为参数 r，则双方的支付函数当满足如下条件：

$$b_{22}/a_{22}=r,\quad b_{12}/a_{12}=b_{21}/a_{21}=-r$$

最后，考虑到支付函数的比例变化对博弈无实质性影响，依旧令 $a_{21}=b_{12}=-1$，于是有

$$r=[(1-p)/p]^2,\quad a_{12}=r^{-1},\quad a_{22}=-r^{-0.5},\quad b_{21}=r,\quad b_{22}=-r^{0.5}$$

将上述支付值代入可得：

$$a_{11}=-r^{-1.5}-r^{-1}-1,\quad b_{11}=-r^{1.5}-r-1$$

显然，a_{11} 与 b_{11} 满足 $b_{11}/a_{11}=r^{1.5}$，比值 $r^{1.5}$ 可分解为车流量比 r 与红绿灯时间比$(1-p)/p=r^{0.5}$ 之乘积。综上，我们已经获得了两个博弈参与人支付函数在战略组合空间上的全部取值，从而得到车辆争道博弈的如下支付矩阵：

$$\mathbf{U}_1=\begin{bmatrix}(-r^{-1.5}-r^{-1}-1,-r^{1.5}-r-1) & (r^{-1},-1)\\ (-1,r) & (-r^{-1.5},r^{-1.5})\end{bmatrix}$$

对于任意给定的 $r\geqslant0$，方程 $r=[(1-p)/p]^2$ 在区间$(0,1)$内的唯一解是 $p=1/(r^{0.5}+1)$，因此，由 U_1 决定的斗鸡博弈的混合战略均衡为$((1/(r^{0.5}+1),r^{0.5}/(r^{0.5}+1)),(r^{0.5}/(r^{0.5}+1),1/(r^{0.5}+1)))$。动态而言，面向东西方向车辆的绿灯时间比重 $p=1/(r^{0.5}+1)$为南北与东西方向车流量比 r 的减函数，这表明随着 r 的增加，面向东西方向车辆的绿灯时间比重 p 是减少的，这符合一般实际情况。特别地，若车流量比为 0，则 $p=1$，若车流量比为 1，则 $p=0.5$。另外，$p=1/(r^{0.5}+1)$是一个凸函数，这意味着 p 的下降速度将随 r 的逐步增加而衰减。

在一定假设基础上，我们构造了一个处于交通监管下的车辆争道博弈，用以规范车辆与车辆之间的争道行为，该博弈给出了具有均衡性质的红绿灯比设置方式。需要引起注意的几点是：

①上述斗鸡博弈混合战略均衡的获得决定于交通监管行为，或者说取决于 a_{11} 与 b_{11} 的强制性赋值，该赋值表征了交通监管的力度。

②考虑到 $r-1$ 与 r 都表征车流量比值，因此，参数 r 的取值可以限制在

区间[0,1]。

③博弈双方的支付都是 r 的函数，而且函数表达形式具有本质上的一致性，因为 $r-1$ 与 r 无本质差异。如此看来，由矩阵 U_1 所定义的博弈具有对称性。

我们所构造的争道监管博弈具有对称性，这一点至关重要，体现了交通监管的公平性。另外，博弈参与人在监管下的支付值仅取决于车流量，完全排除了潜在的交通意外对支付函数的影响，这要求有完善的交通保险制度提供支持。当然，支付函数的客观性有其优点，在一定程度上确保了交通监管的实际可操作性。

(5)规范行人—车辆争道行为的红绿灯设置

现考虑规范行人与车辆争道行为的红绿灯时间设置问题。此时的交通灯乃为方便行人横过公路而设置，行人与车辆在监管下的对局如表 7-5 所示。

表 7-5　监管下的行人—车辆博弈

行人＼车辆	绿灯	红灯
绿灯	a_{11},b_{11}	a_{12},b_{12}
红灯	a_{21},b_{21}	a_{22},b_{22}

与十字路口车辆争道博弈相区别，作为博弈参与人的行人与车辆在支付函数方面未必具有可比性，需要施加更为严格的假设。我们沿用前方假设 1 与 3，并修正假设 2 为：车辆与行人单位时间收益之比为常数，并设该参数为 γ。与前方类似，可得到如下支付矩阵：

$$\mathbf{U}_2=\begin{bmatrix} -\gamma^{-1.5}(-\gamma^{-1}-1,-\gamma^{1.5}-\gamma-1) & (\gamma^{-1},-1) \\ (-1,\gamma) & (-\gamma^{-0.5},\gamma^{-0.5}) \end{bmatrix}$$

显然，a_{11} 随着 γ 的增加而增加，b_{11} 则随着 γ 的增加而下降，这是人性化监管与有效性监管的基本要求。监管的人性化表现在交通意外问责方面，行人或车辆违纪极有可能导致交通事故，高受益者理所当然承担更多的事故责任，因此，γ 的增加意味着车辆应该承担的道义成本在增加，使得 b_{11} 取值上扬。监管有效性指的是惩戒力度的有效性，就一般情况而言，随着车辆行人收益比 γ 的增加，车辆等待的机会成本相应增加，则车辆违规的潜在可能性上升，需要施以严厉的惩戒。反之，γ 的增加意味着行人等待成本下降，其违规的可能性相应减少，则针对行人违纪的惩罚可酌情放松，因此，a_{11} 与 b_{11} 的变动是反向的。

如前所述，红绿灯时间比值 $p=1/(1+\gamma^{0.5})$ 最终决定于车辆与行人单位时间收益比 γ，该数值难以直接获得。另外，实际操作时还需要关注车辆、行

人流量等直接影响 γ 取值的现场数据。为简单起见，取车辆行人收益比 γ 为资本劳动收益比，则 p 值可得以确定。比如，中国2004年劳动收入占GDP比重约为0.42，资本劳动收益比 $\gamma=1.38$，可求得 $p=0.46$，因此，对一个具代表性的交通灯而言，若红绿灯周期为1 min，则面向行人的绿灯时长约28 s，而红灯时长约32 s。

(6)结论

基于博弈论，本项研究考察红绿灯所行使的如下3项职能：

每一个给定的红绿灯状态都对应一个纯战略纳什均衡；

红绿灯状态的变化对应纯战略纳什均衡的变化；

绿信比对应混合战略纳什均衡。

以斗鸡博弈描述争道问题，并且在时间价值基础上构造了争道斗鸡博弈的支付函数，利用博弈的混合战略纳什均衡确定了符合博弈均衡特征的绿信比，该均衡绿信比可表示为不同方向车流或人流比的一个具体函数。若行人与车辆满足经济人基本假设，则在相应的监管制度的保障下，按照这一程式确定的绿信比符合博弈均衡特征，可增加行人或车辆遵守交通规章的自觉性，必将有利于城市交通状况的进一步改善。

7.4.4 企业竞争博弈

本节通过建立两个企业之间竞争的博弈模型，从完全信息静态博弈、完全信息动态博弈和不完全信息博弈三个角度分析论证了博弈的结果，并在此结果的基础上提出有利于企业合作竞争的相关政策建议。

(1)完全信息博弈分析

①完全信息静态博弈分析。假设存在甲、乙两个企业，双方的产品、企业规模和所处环境等因素相似，并且均希望通过竞争使本企业在市场上占据领先地位。现在市场上有一笔大的订单，但企业必须给出优惠的条件才能吸引对方，企业双方想联合得到这笔订单，每个企业都有合作和不合作两种选择。现构建一个博弈模型。在联合获取这笔订单的过程中，如果双方均选择合作，即双方共同出资联合竞争这笔订单，制定并执行统一的价格，则甲、乙双方均可获得10个单位的收益。如果一个企业选择不合作，则该企业需为其优惠条件付出3个单位的收益。那么，如果两个企业都选择不合作，则双方获得的收益均为7。如果一个企业选择合作，而另一个企业选择不合作，则选择不合作的企业能够提供更为优惠的条件，从而将对方50%的收益吸引过来，因此，不合作方可获得 $7+10\times50\%=12$ 个单位的收益，合作方的收益为 $10-10\times50\%=5$ 个单位。根据上面所说博弈条件，可列出双方博弈的支付矩阵，

见表 7-6。

表 7-6 甲、乙企业博弈的均衡情况

乙 \ 甲	合作	不合作
合作	10,10	5,12
不合作	12,5	7,7

假设甲、乙两个企业均为理性人，那么它们都会按照实现其利益最大化的原则进行选择。在博弈的过程中，如果让乙先进行选择，无论甲选择合作或不合作，乙都是选择不合作带来的收益最大。同样当甲企业进行选择时也会选择不合作。那么对甲乙两个企业来说，不管对方的选择是什么，自己的占优策略就是选择不合作，最终的纳什均衡是(不合作，不合作)，收益组合是(7,7)。在这种状态下，双方都没有动力做进一步改变。然而对整体而言，(合作，合作)这个博弈选择才是对双方来说都最有利的均衡解。但是，在理性的假设下，博弈的一方不能控制另一方的决策，企业双方都会有选择不合作的动力，因此它们都从各自短期利益出发，走进了“囚徒困境”。

在竞争激烈的市场环境下，各企业都是从各自的利益出发，而不是从整体利益考虑问题，在缺少有效的外部约束的情况下，博弈最后形成的结果只能是“囚徒困境”。这种结果不仅损害了整体的利益，使得市场秩序被破坏，社会效用被降低，而且这样不合作的后果可能导致一种毁灭性的恶性竞争局面。

②完全信息动态博弈分析。下面我们从动态的角度来对企业竞争行为进行分析。假设甲、乙企业仍符合理性人的条件，在第一轮博弈结束后，双方都选择不合作，则意味着双方企业都将会提供更加优惠的条件来竞争这笔订单，此时，如果双方给出的优惠条件相同或者接近，使得对方无法抉择，那么甲、乙两企业为了竞争这笔订单势必将提供更加优惠的条件。但是无论是哪一个企业先行动，最后出现的结果还是(不合作，不合作)。

在第一轮博弈的基础上，如果企业双方都选择合作，表示双方都接受第一轮最后对方提供的优惠条件，都不改变，则都可获得 7 个单位的收益。如果都选择不合作，则由于它们均提供了更为优惠的条件吸引订单，因此它们的收益均为 4 个单位。如果一方选择合作，一方选择不合作，那么选择不合作的企业必定先给出更为优惠的条件与另一方竞争，假定可以吸收对方 50%的收益，则不合作的企业将会获得 $4+7\times50\%=7.5$ 个单位的收益，选择合作的企业获得的收益为 $7-7\times50\%=3.5$ 个单位。在第一轮博弈的基础上，第二轮的博弈中甲、乙的均衡情况如表 7-7。

表 7-7　甲、乙企业博弈的均衡情况

乙＼甲	合作	不合作
合作	7,7	3.5,7.5
不合作	7.5,3.5	4,4

第二轮博弈最后形成的均衡结果还是(不合作,不合作),这又出现了新的“囚徒困境”。新一轮的博弈结束,双方获得的收益将更少,从而使得双方本该收益最大化的合作行为变为同时降低收益的竞争行为,最终导致资源的浪费和效率的降低,这就是企业竞争的结果。如果继续照此进行下去,双方的收益还会进一步减少。

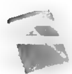
但是,无限次的重复博弈竞争可以在一定程度上改变这种状况,因为多次选择不合作的反悔行为会对企业声誉造成一定程度的影响,另一方面,如果双方企业在博弈中的收益一再减少,也必定会通过协议商谈等方式改变这种状况。然而,在完全信息情况下,实现合作需要双方具有无限次博弈机会以及足够的耐心。而且博弈机会从来都是有限资源,不可能博弈无限多次。

(2)不完全信息博弈分析

在现实生活中,各企业不可能完全理性,企业间的信息不完全或不对称是常态,因此,企业间的竞争更符合不完全信息情况下的有限次重复博弈的动态选择。下面证明在不完全信息条件下,只要博弈重复次数足够多,在有限次博弈中可以走出“囚徒困境”,合作行为会出现。

在上面完全信息静态博弈的模型下,假设企业双方均为理性人,那么它们最后的占优策略为(不合作,不合作)。假设一个企业是理性的,另一个企业是非理性的,非理性的企业对自己的声誉看得非常重,在第一阶段它必然会选择合作,在以后的阶段,它将会参考对方企业第一阶段的选择,如果对方合作它将选择合作,如果对方不合作它也将选择不合作;而理性的企业通过博弈分析,不管对方如何选择,它都会选择不合作。因此在此假设下,博弈的均衡为(不合作,合作)。假设企业双方均为非理性人,那么它们不管对方怎么选择都将会选择合作,所以这次博弈的均衡结果是(合作,合作)。

以上分析的是不完全信息条件下的一次性博弈,接下来分析不完全信息情况下的重复博弈。假设企业甲、乙分别有理性和不理性两种类型,双方都只知道自己的类型,且知道对方属于非理性的概率是 p,属于理性的概率是 $1-p$。非理性企业不能原谅对方任何反悔的行为,因此如果对方在 t_0 阶段选择不合作,它将会从 t_0+1 阶段开始选择不合作,直到整个博弈结束;而理性企业可以采取任何认为对自己有利的策略。假定双方是在一轮博弈结束、观



察到这轮博弈的结果后进行下一轮博弈的，而且博弈重复的次数 t 足够多；博弈双方的收益是每一轮博弈的收益贴现值之和，设贴现因子为 $\theta(0<\theta<1)$。

首先，假设企业甲为理性的，它在第一阶段可以选择合作或不合作。如果甲选择不合作，那么从第二阶段开始，不管企业乙是理性的还是非理性的都会选择不合作，因为如果乙是理性的，选择不合作收益更大；如果乙是非理性的，选择不合作是为了报复甲在第一阶段的背信弃义。那么从第二阶段开始直到所有回合结束，企业甲也都会选择不合作。即两个企业都将选择不合作直到最后。根据这个战略，理性企业甲的第一阶段合作期望收益是：

$$12+7\theta+7\theta^2+\cdots+7\theta^{(t-1)}=5+7(1-\theta^t)/(1-\theta)$$

其次，考虑企业甲的另一种战略，假设企业甲会首先选择合作，直到对方在某阶段(t_0)选择不合作，然后从 t_0+1 开始选择不合作直到博弈的所有回合结束。那么，如果企业乙是非理性的(发生的概率是 p)，企业乙会一直选择合作，则企业甲也会一直选择合作。这种情况下，企业甲得到的收益是：$10+10\theta+10\theta^2+\cdots+10\theta^{(t-1)}=10(1-\theta^t)/(1-\theta)$。如果企业乙是理性的(发生的概率是 $1-p$)，不管他开始时选择合作还是不合作，从 t_0+1 开始他将选择不合作直到最后，那么，企业甲得到的最小收益是：

$$5+7\theta+7\theta^2+\cdots+7\theta^{(t-1)}=7(1-\theta^t)/(1-\theta)-2$$

因此，非理性企业甲的最小期望收益是：

$$10p(1-\theta^t)/(1-\theta)+(1-p)[7(1-\theta^t)/(1-\theta)-2]$$

那么，如果满足 $5+7(1-\theta^t)/(1-\theta)>10p(1-\theta^t)/(1-\theta)+(1-p)[7(1-\theta^t)/(1-\theta)-2]$，即 $t<\ln[1-(7-2p)(1-\theta)/(3p)]/\ln\theta$，表示企业甲的第二种战略肯定优于第一种战略。也就是说，只要 t 足够大，企业甲从一开始就选择不合作就不是最优战略。由此说明，只要双方重复博弈的次数足够多，当博弈次数超过某一临界值时，在第一阶段选择不合作对理性企业而言不是最优战略，即此时企业双方选择合作才是最优战略。

思考题

1. 厂家价格大战的结局应该是一个“纳什均衡”，而且价格战的结果是谁都没钱赚。因为博弈多方的利润正好是零。竞争的结果是稳定的，即是一个“纳什均衡”。这个结果可能对消费者是有利的，对厂商而言是灾难性的。事实上，价格战始终被许多企业惯用。通过反复地发起价格战，格兰仕成为微波炉霸主，长虹曾一度主导整个彩电市场。试分析其中的奥妙。

2. 任何一个国家在国际贸易中都面临着保持贸易自由与实行贸易保护主义的两难选择。贸易自由与壁垒问题，也是一个“纳什均衡”，这个均衡是

贸易双方采取不合作博弈的策略,结果使双方因贸易战受到损害。试分析最近两年中、日、韩三国贸易战。

3. 枪手博弈。3人对枪自决,甲乙丙枪法优劣递减,每人先选择1人射击。详细分析讨论各种可能结果,谁最有可能活下来?

4. 钓鱼岛争端——大国间的博弈。2012年由日本政府"购岛"引发的钓鱼岛争端升级,钓鱼岛的战略地位非常重要,美国明显倾向于日本。围绕钓鱼岛争端,试分析中、日、美三国的政治博弈。

第 8 章

图论与网络分析

很多人有饭后散步的习惯，由此自然联想到哥尼斯堡七桥问题：哥尼斯堡（现名加里宁格勒）是欧洲一个城市，Pregei 河把该城分成两部分，河中有两个小岛，18 世纪时，河两边及小岛之间共有 7 座桥，当时人们提出这样的问题：有没有办法从某处（如 A）出发，经过各桥一次且仅一次最后回到原地（见图 8-1）。最后，数学家 Euler 在 1736 年巧妙地给出了这个问题的答案，并因此奠定了图论的基础，Euler 把 A、B、C、D 四块陆地分别收缩成四个顶点，把桥表示成连接对应顶点之间的边，问题转化为从任意一点出发，能不能经过各边一次且仅一次，最后返回该点（见图 8-2）。这就是著名的 Euler 问题，他将此研究结果于 1736 年发表，被公认为是有关图论的第一篇论文，由此拉开了对图论的研究，Euler 本人也被尊崇为图论和拓扑学之父。

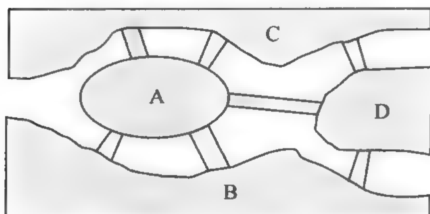


图 8-1　哥尼斯堡七桥问题

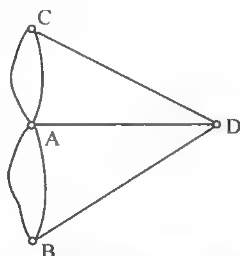


图 8-2　哥尼斯堡七桥问题的图表示

图论是专门研究图的理论的一门数学分支，属于离散数学范畴，与运筹学有交叉，它有 200 多年历史，大体可划分为三个阶段。第一阶段是从 18 世纪中叶到 19 世纪中叶，处于萌芽阶段，多数问题因游戏而产生（如上所谓的 Euler 七桥问题，即一笔画问题）。第二阶段是从 19 世纪中叶到 20 世纪中叶，这时图论问题大量出现，如 Hamilton 问题、地图染色的四色问题以及图的可平面性问题等。这时，也出现了用图的理论来解决实际问题：如 Cayley 把树应用于化学领域，Kirchhoff 用树去研究电网络等。第三阶段是 20 世纪中叶以后，在图论理论研究的同时，出现了图论的大量应用，人们对物理、化学、通讯科学、建筑学、生物遗传学、心理学、经济学、生产管理、军事、交通、运输、计算机网络等方面提出的许多实际问题应用图论的理论来解决。由于大型计

算机的使用使大规模问题的求解成为可能，同时 Ford 和 Fulkerson 建立网络流理论使图论与线性规划、动态规划等优化理论和方法相互渗透，从而进一步促进了图论在实际问题中的应用。究竟图论包含哪些内容，下面我们进行简单讨论。

8.1 基础知识简介

先看一个问题：下列四个图形中哪些可以一笔画（见图 8-3）？

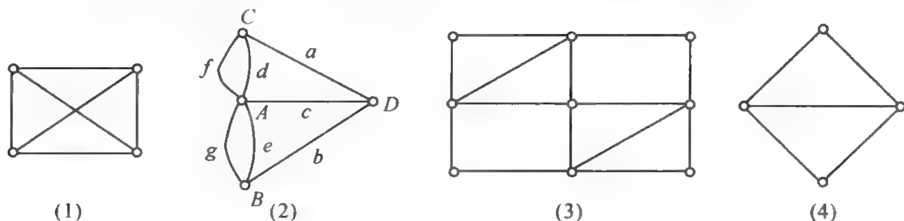


图 8-3 讨论一笔画的图例

当然，图(1)(2)不能一笔画，(3)(4) 能一笔画，且(4)中笔又能回到出发点。这是为什么？为讨论问题的方便，我们需要将以上的图进行数学表示。

8.1.1 基本定义

定义 1 一个图 G 是由点和连线组成（连线可带箭头，称为弧，也可不带箭头，称为边）。

一个图的点构成点集 $V(G)$，一个图的连线构成边集 $E(G)$，它们构成的二元组称为图 G，记作 $G=(V(G),E(G))$，简记为 $G=(V,E)$（注：边与顶点之间的连接是一个对应关系，为了直观，通常用图的图示表达），其中 V 中的元素叫做顶点（简称为点），V 表示图 G 的顶点（或点）集合，E 中的元素叫做边，E 表示图 G 的边集合。

如果一个图是由点和边所构成的，则称其为无向图，记作 $G=(V,E)$，连接点的边记作 v_iv_j 或者 a 等。如图 8-3 中，若(2)的图记为 $G=(V,E)$，则

$$V=\{A,B,C,D\},E=\{a,b,c,d,e,f,g\},$$

其中 $a=CD,b=BD,c=AD,d=CA,e=AB,f=CA,g=AB$。

如果一个图是由点和弧所构成的，那么称之为有向图，记作 $D=(V, A)$，其中 V 表示有向图 D 的顶点集合，A 表示有向图 D 的弧集合。一条方向从 v_i 指向 v_j 的弧，记作 a 或者 (v_i,v_j)。当弧 $a=(v_i,v_j)$时，称 v_i 为 a 的尾，v_j 为 a 的头，并称弧 a 为 v_i 的出弧，为 v_j 的入弧。如图 8-4 中，若将其有向图记

为 $G=(V,A)$，则

$$V=\{A,B,C,D\},E=\{a,b,c,d,e,f\},$$

其中 $a=(D,C),b=(B,C),c=(B,D),d=(B,A),e=(A,B),f=(A,D)$。

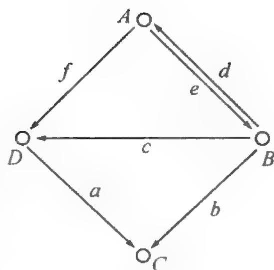


图 8-4

定义 2 图的一条边的两个端点是相同的，那么称这条边是环。如果两个端点之间有两条或两条以上的边，那么称它们为多重边。

定义 3 一个无环、无多重边的图称为简单图，每一对顶点间都有边相连的无向简单图称为完全图。n 个顶点的完全图记为 K_n。

定义 4 无向图中顶点 v 为端点的边的个数称为点 v 的度数（简称度），记作 $d(V)$。度数为零的顶点称为弧立点，度数为 1 的顶点称为悬挂点。悬挂点的关联边称为悬挂边。度数为奇数的点称为奇点，度数为偶数的点称为偶点。

定义 5 有向图中，以 v_i 为始点的边数称为点 v_i 的出次（或出度），用 $d^+(v_i)$ 表示；以 v_i 为终点的边数称为点 v_i 的入次（或入度），用 $d^-(v_i)$ 表示。

用 $\Delta^+(G),\Delta^-(G),\delta^+(G),\delta^-(G)$（简记为 $\Delta^+,\Delta^-,\delta^+,\delta^-$）分别表示图 G 的最大出次和入次、最小出次和入次。

8.1.2 基本结论

定理 1 所有顶点的度数之和等于总边数的 2 倍。

定理 2 在任一图中，奇点的个数必为偶数。

定理 3 所有顶点的入次之和等于所有顶点的出次之和，等于图的总弧数。

其余的定义和结论在具体问题的建模讨论中再给出。

8.2 哥尼斯堡七桥问题与欧拉图

8.2.1 基本定义

在一笔画的讨论中，我们需要将无向图的边按一定顺序排列，为此我们给出以下定义。

定义 6 对图 $G(V(G),E(G))$ 来说，当边 $e_k=v_iv_j$ 时，称 v_i,v_j 为边 e_k 的端点，并称 v_j 与 v_i 相邻；边 e_k 称为与顶点 v_i,v_j 关联。如果某两条边至少有一个公共端点，则称这两条边在图 G 中相邻。

定义7 给定一个图 $G=(V,E)$ 或有向图 $D=(V,A)$，在 V 中指定两个点，一个称为始点(或发点)，记作 v_0，一个称为终点(或收点)，记作 v_k，令 $W=v_0e_1v_1e_2\cdots e_kv_k$，其中 $e_1\in E(G)$，$1\leqslant i\leqslant k$，$v_j\in V(G)$，$0\leqslant j\leqslant k$，e_i 与 v_{i-1}，v_i 关联，称 W 是图 G 的一条道路(或途径)，k 为路长，顶点 v_0 和 v_k 也可分别称为 W 的起点和终点，而 $v_1,v_2,\cdots,v_{k-1}$ 称为它的内部顶点或中间点。若道路 W 的边互不相同，则 W 称为迹。若道路 W 的顶点互不相同，则 W 称为轨(或路)。

定义8 称一条道路是闭的(或称为回路)，如果它有正的长度且起点和终点相同，起点和终点重合的迹叫做闭迹，起点和终点重合的轨叫做圈。

定义9 若图 G 的两个顶点 u，v 间存在道路，则称 u 和 v 连通(connected)。u，v 间的最短轨的长叫做 u，v 间的距离，记作 $d(u,v)$。若图 G 的任两顶点均连通，则称 G 是连通图。

8.2.2 欧拉回路与欧拉图

定义10 连通图 G 中，若存在一条道路，经过每边一次且仅一次，则称这条路为欧拉道路。若存在一条回路，经过每边一次且仅一次，则称这条回路为欧拉回路(或 Euler 环游)，具有 Euler 环游的图称为欧拉图。

即经过 G 的每条边的迹叫做 G 的 Euler 迹；闭的 Euler 迹叫做 Euler 回路或 E 回路；含 Euler 回路的图叫做 Euler 图。直观地讲，Euler 图就是从一顶点出发每边恰好通过一次能回到出发点的那种图，即不重复地行遍所有的边再回到出发点的图。

定理4 一个多重连通图 G 是欧拉图的充分必要条件是 G 中无奇点，一个多重连通图 G 有欧拉道路的充分必要条件是 G 有且仅有两个奇点。

8.2.3 哥尼斯堡七桥问题的解答

定理4给出了哥尼斯堡七桥问题否定的回答。但是对欧拉图如何找到 Euler 环游的问题以及如何找到欧拉道路的问题，有以下算法。

(1)欧拉图中找 Euler 环游的算法——求 Euler 回路的 Fleury 算法

1921 年，Fleury 给出下面的求 Euler 回路的算法(直观地讲，他是“过河拆桥，尽量不走独木桥”的方法)。

Fleury 算法：

① $\forall v_0\in V(G)$，令 $W_0=v_0$。

②假设迹 $W_i=v_0e_1v_1\cdots e_iv_i$ 已经选定，那么按下述方法从 $E-\{e_1,\cdots e_i\}$ 中选取边 e_{i+1}：

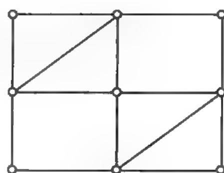
e_{i+1}和 v_i 相关联；

除非没有别的边可选择，否则 e_{i+1} 不是 $G_i = G - \{e_1, \cdots, e_i\}$ 的割边（所谓割边是一条删除后使连通图不再连通的边）。

③当第 2 步不能再执行时，算法停止。

(2)找欧拉道路算法

由于简单连通图 G 恰好有两个奇点，可将 G 的两奇点连一条边后得到的图设为 H，则图 H 为欧拉图，在 H 中取 G 的任一奇点作为 v_0，用 Fleury 算法求出 H 的 Euler 回路，删去新加的边即得图 G 的欧拉道路。


图 8-5

例 8.1 用 Fleury 算法求图 8-5 的 Euler 回路。

解：求法见图 8-6，其中 v_0 取为 A 点。

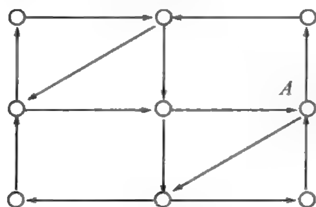
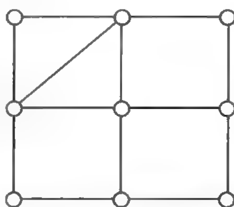


图 8-6


图 8-7

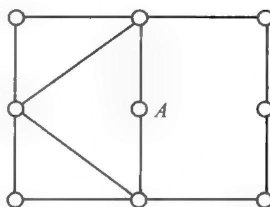


图 8-8

动脑筋：

(1)求图 8-7 的欧拉道路。

(2)请你为服务图 8-8 所示街区的邮递员设计一个送信方案（邮局在 A 点处）。

8.3 匹配问题

前文已提及中国邮递员问题，要解决这一问题，需要用到最短路算法和完美对集，而完美对集有着广泛的应用。例如用以下定理 6 可证明：每个男孩都能娶到一个他所认识的女孩的充要条件是任意 k 个男孩都至少认识 k 个女孩。现在来讨论求完美对集的算法，这属于匹配问题。

8.3.1 基本定义

定义 11 若 $M \subset E(G)$，$\forall e_i, e_j \in M$，e_i 与 e_j 无公共端点（$i \neq j$），则称 M 为图 G 的一个对集；M 中的一条边的两个端点叫做在对集 M 中相配；M 中的端点称为被 M 许配；G 中每个顶点皆被 M 许配时，M 称为完美对集；G 中已

无使$|M'|>|M|$的对集M'，则M称为最大对集；若G中有一轨，其边交替地在对集M内外出现，则称此轨为M的交错轨，交错轨的起止顶点都未被M许配时，此交错轨称为M的可增广轨。

例如：一个图的最大匹配和完美匹配，见图 8-9（粗线表示）。

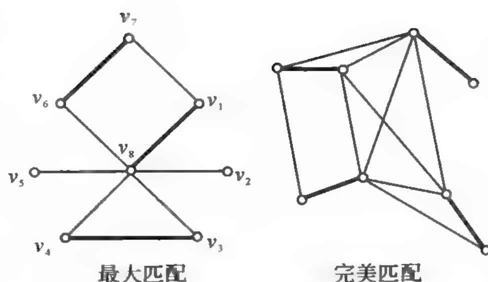


图 8-9

若把可增广轨上在M外的边纳入对集，把M内的边从对集中删除，则被许配的顶点数增加 2，对集中的“对”增加一个。

定义 12　若$V(G)=X\cup Y, X\cap Y=\varnothing, |X||Y|\neq 0$（这里$|X|$表示集合$X$中的元素个数），$X$中无相邻顶点对，$Y$中亦然，则称$G$为二分图；特别地，若$\forall_x\in X, \forall_y\in Y$，则$xy\in E(G)$，则称$G$为完全二分图，记成$K_{|X|,|Y|}$。

8.3.2　基本结论

1957 年，贝尔热（Berge）得到最大对集的充要条件：

定理 5　M是图G中的最大对集当且仅当G中无M可增广轨。

1935 年，霍尔（Hall）得到下面的许配定理：

定理 6　G为二分图，X与Y是顶点集的划分，G中存在把X中顶点皆许配的对集的充要条件是：$\forall S\subset X$，则$|N(S)|\geqslant|S|$，其中$N(S)$是S中顶点的邻集。

由上述定理可以得出：

推论 3.1：若G是$k(k>0)$正则二分图，则G有完美对集。

所谓k正则图，即每顶点皆k度的图。由此推论得出下面的婚配定理：

定理 7　每个姑娘都结识$k(k\geqslant 1)$位小伙子，每个小伙子都结识k位姑娘，则每位姑娘都能和她认识的一个小伙子结婚，并且每位小伙子也能和他认识的一个姑娘结婚。

例 8.2　人员分派问题：工作人员$x_1, x_2, \cdots, x_n$去做n件工作$y_1, y_2, \cdots, y_n$，每人适合做其中一件或几件，问能否每人都有一份适合的工作？如果不能，最多几人可以有适合的工作？

这个问题的数学模型是：G是二分图，顶点集划分为$V(G)=X\cup Y$，$X=\{x_1,\cdots,x_n\}$，$x_i\in X$，$Y=\{y_1,\cdots,y_n\}$，$y_i\in Y$，当且仅当x_i适合做工作y_i时，$x_iy_i\in E(G)$，求G中的最大对集。

解决这类问题可以利用1965年埃德门兹(Edmonds)提出的匈牙利算法。

求 G 中的完美对集的匈牙利算法：

(1)从 G 中任意取定一个初始对集 M。

(2)若 M 把 X 中的顶点皆许配，停止，M 即完美对集；否则取 X 中未被 M 许配的一顶点 u，记 $S=\{u\}, T=\varnothing$。

(3)若 $N(S)=T$，停止，无完美对集；否则取 $y\in N(S)-T$。

(4)若 y 是被 M 许配的，设 $yz\in M, S=S\cup\{z\}, T=T\cup\{y\}$，转(3)；否则，取可增广轨 $P(u,y)$，令 $M=(M-E(P))\cup(E(P)-M)$，转(2)。

把以上算法稍加修改就能够用来求二分图的最大对集。

例 8.3 如图8-10中的(1)所示，$V_1=\{x_1,x_2,x_3,x_4,x_5\}$，$V_2=\{y_1,y_2,y_3,y_4,y_5\}$，试求图 G 的最大匹配。

解：任取初始匹配 $M=\{x_2y_2, x_3y_3, x_5y_5\}$，如图8-10的(1)中虚线所示。解题过程如下表。

表 8-1 算法过程

M	x	S	T	$N(S)$	$y\in N(S)-T$	$\{y,u\}\in M$	P
$\{x_2y_2,x_3y_3,x_5y_5\}$	x_1	$\{x_1\}$	$\varnothing$	$\{y_2,y_3\}$	y_2 饱和	$\{y_2,x_2\}$	
		$\{x_1,x_2\}$	$\{y_2\}$	$\{y_1,y_2,y_3,y_4,y_5\}$	y_1 非饱和		(x_1y_2,x_2y_1)
$\{x_1y_2,x_2y_1,x_3y_3,x_5y_5\}$	x_4	$\{x_4\}$	$\varnothing$	$\{y_2,x_3\}$	y_2 饱和	$\{y_2,x_1\}$	
		$\{x_4,x_1\}$	$\{y_2\}$	$\{y_2,y_3\}$	y_3 饱和	$\{y_3,x_3\}$	
		$\{x_4,x_1,x_3\}$	$\{y_2,y_3\}$	$\{y_2,y_3\}$	$N(S)=T$，停止		

所以，$M=\{x_1y_2,x_2y_1,x_3y_3,x_5y_5\}$ 即为图 G 的最大匹配，如图8-10的(2)中虚线所示。

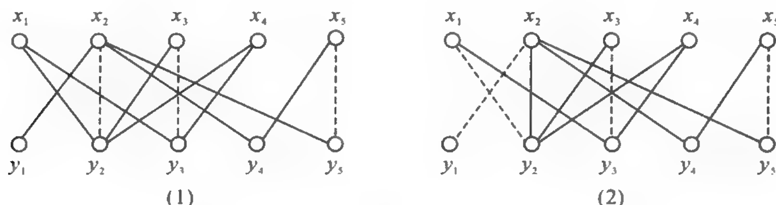


图 8-10 求解图

解决这个问题可以用库恩—曼克莱斯(Kuhn-Munkres)算法。为此，我们要引入可行顶点标号与相等子图的概念。

定义 13 若映射 $l:V(G)\to R$，满足 $\forall x\in X,y\in Y$，

$$l(x)+l(y)\geqslant w(x,y),$$

则称 l 是二分图 G 的可行顶点标号。令

$$E_l=\{xy\mid xy\in E(G),l(x)+l(y)=w(xy)\},$$

称以 E_l 为边集的 G 的生成子图为相等子图，记作 G_l。

二分图的可行顶点标号是存在的。例如

$$l(x)=\max_{y\in Y}w(xy),\quad x\in X;$$

$$l(y)=0,\qquad y\in Y。$$

定理 8 G_l 的完美对集即为 G 的权最大的完美对集。

Kuhn-Munkres 算法：

(1)选定初始可行顶点标号 l，确定 G_l，在 G_l 中选取一个对集 M。

(2)若 X 中顶点皆被 M 许配，停止，M 即 G 的权最大的完美对集；否则，取 G_l 中未被 M 许配的顶点 u，令 $S=\{u\}$，$T=\varnothing$。

(3)若 $N_{G_l}(S)\supset T$，转(4)；若 $NG_l(S)=T$，取

$$\alpha_1=\min_{x\in S,y\not\subset T}\{l(x)+l(y)-w(xy)\},$$

$$\bar{l}(v)=\begin{cases}l(v)-\alpha_1,v\in S\\ l(v)+\alpha_1,v\in T,\\ l(v),\text{其他}\end{cases}$$

$$l=\bar{l},\ G_l=G_{\bar{l}}。$$

(4)选 $N_{G_l}(S)-T$ 中一顶点 y，若 y 已被 M 许配，且 $yz\in M$，则 $S=S\cup\{z\}$，$T=T\cup\{y\}$，转(3)；否则 G_l 中一个 M 可增广轨 $P(u,y)$，令

$$M=(M-E(P))\cup(E(P)-M),$$

转(2)。

其中 $N_{G_l}(S)$是 S 的相邻顶点集。

以上算法稍加修改，可得权最小的完美对集(见孙惠泉编著《图论及其应用》，科学出版社 2006 年版)。

例 8.4 已知完全二分图 $K_{5,5}$，其中 $V_1=\{x_1,x_2,x_3,x_4,x_5\}$，$V_2=\{y_1,y_2,y_3,y_4,y_5\}$，且 $K_{5,5}$ 的权矩阵为 A(见图 8-11)，求 $K_{5,5}$ 的权最大的最优匹配。

$$A=\begin{pmatrix}3&5&5&4&1\\2&2&0&2&2\\2&4&4&1&0\\0&1&1&0&0\\1&2&1&3&3\end{pmatrix}$$

图 8-11

解：(1)取可行顶点标 l 如下：

$l(y_1)=l(y_2)=l(y_3)=l(y_4)=l(y_5)=0$

$l(x_1)=\max(3,5,5,4,1)=5$

$l(x_2)=\max(2,2,0,2,2)=2$

$l(x_3)=\max(2,4,4,1,0)=4$

$l(x_4)=\max(0,1,1,0,0)=1$

$l(x_5)=\max(1,2,1,3,3)=3$

(2)取 G_l 及 G_l 是匹配，如图 8-12 所示（虚线）。由于 $W(G-x_2)=3$，故 G_l 中无完美匹配，需修改顶标。

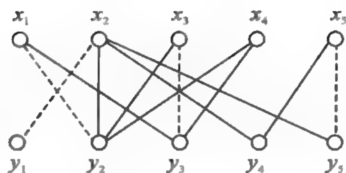


图 8-12

(3) $x=x_4$，得 $S=\{x_4,x_1,x_3\}$，$T=\{y_2,y_3\}$，$N_{G_l}(S)=T$，于是

$$\alpha_1=\min\{\ l(x)+l(y)-(x,y)\,|\,x\in S,y\in V_2-T\}=1$$

因而 x_1,x_2,x_3,x_4,x_5 的顶标分别为 4,2,3,0,3；y_1,y_2,y_3,y_4,y_5 的顶标分别为 0,1,1,0,0。

(4)用修改后的顶标 l，得 G_l 及 G_l 的一个匹配，如图 8-13 所示（虚线）。此匹配即 $K_{5,5}$ 的最优匹配。其总权为 2+4+1+4+3=14。

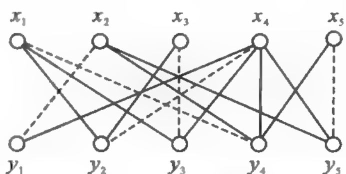


图 8-13　G_l 及它的完美匹配

说明：图 G 的最优匹配未必唯一。

除了最优匹配外，还有倾向度分派意义下的稳定匹配问题（有关内容参见孙惠泉编著《图论及其应用》，科学出版社 2006 年版）。

8.4　中国邮递员问题

8.4.1　最短路问题

求最短路已有成熟的算法，如迪克斯特拉(Dijkstra)算法，其基本思想是按距 u_0 从近到远为顺序，依次求得 u_0 到 G 的各顶点的最短路和距离，直至 v_0（或直至 G 的所有顶点），算法结束。为避免重复并保留每一步的计算信息，通常采用标号算法。

(1) 令 $l(u_0)=0$，对 $v\neq u_0$，令 $l(v)=\infty$，$S_0=\{u_0\}$，$i=0$。

(2) 对每个 $v\in\bar{S}_i(\bar{S}_i=V/S_i)$，用 $\min\limits_{u\in S_1}\{l(v),l(u)+w(uv)\}$ 代替 $l(v)$。计算 $\min\limits_{v\in \bar{S}_i}\{l(v)\}$，把达到这个最小值的一个顶点记为 u_{i+1}，令 $S_{i+1}=S_i\cup\{u_{i+1}\}$。

(3) 若 $i=|V|-1$，停止；若 $i<|V|-1$，用 $i+1$ 代替 i，转(2)。

算法结束时，从 u_0 到各顶点 v 的距离由 v 的最后一次的标号 $l(v)$ 给出。在 v 进入 S_i 之前的标号 $l(v)$ 叫 T 标号，v 进入 S_i 时的标号 $l(v)$ 叫 P 标号。算法就是不断修改各顶点的 T 标号，直至获得 P 标号。若在算法运行过程中，将每一顶点获得 P 标号所由来的边在图上标明，则算法结束时，u_0 至各顶点的最短路也在图上标示出来了。

例 8.5 某公司在 6 个城市 $c_1, c_2, \cdots, c_6$ 中有分公司，从 c_i 到 c_j 的直接航程票价记在图 8-14 描述的矩阵的 (i,j) 位置上（∞ 表示无直接航路），请帮助该公司设计一张城市 c_1 到其他城市间的票价最便宜的路线图。

$$\begin{pmatrix} 0 & 50 & \infty & 40 & 25 & 10 \\ 50 & 0 & 15 & 20 & \infty & 25 \\ \infty & 15 & 0 & 10 & 20 & \infty \\ 40 & 20 & 10 & 0 & 10 & 25 \\ 25 & \infty & 20 & 10 & 0 & 55 \\ 10 & 25 & \infty & 25 & 55 & 0 \end{pmatrix}$$

图 8-14　已知矩阵

用矩阵 $a_{n\times n}$（n 为顶点个数）存放各边权的邻接矩阵，行向量 pb、$index_1$、$index_2$、d 分别用来存放 P 标号信息、标号顶点顺序、标号顶点索引、最短通路的值。其中分量

$$pb(i)=\begin{cases}1, & \text{当第 } i \text{ 顶点已标号}\\ 0, & \text{当第 } i \text{ 顶点未标号}\end{cases}$$

$index_2(i)$ 存放始点到第 i 点最短通路中第 i 顶点前一顶点的序号；$d(i)$ 存放由始点到第 x_i 点最短通路的值。

求第一个城市到其他城市的最短路径的 Matlab 程序如下（结果自己求出）：

```
clear;
clc;
M=10000;
a(1,:)=[0,50,M,40,25,10];
a(2,:)=[zeros(1,2),15,20,M,25];
a(3,:)=[zeros(1,3),10,20,M];
a(4,:)=[zeros(1,4),10,25];
a(5,:)=[zeros(1,5),55];
a(6,:)=zeros(1,6);
a=a+a';
pb(1:length(a))=0;pb(1)=1; index1=1; index2=ones(1,length(a));
d(1:length(a))=M;d(1)=0;temp=1;
while sum(pb)<length(a)
    tb=find(pb==0);
    d(tb)=min(d(tb),d(temp)+a(temp,tb));
    tmpb=find(d(tb)==min(d(tb)));
    temp=tb(tmpb(1));
    pb(temp)=1;
    index1=[index1,temp];
    index=index1(find(d(index1)==d(temp)-a(temp,index1)));
    if length(index)>=2
        index=index(1);
    end
    index2(temp)=index;
```

```
end
d, index1, index2
```

8.4.2 中国邮递员问题

一位邮递员从邮局选好邮件去投递，然后返回邮局，他必须经过自己负责投递的每条街道至少一次，为他设计一条投递路线，使得他行程最短。

定义 14 若对于图 $G=(V,E)$ 和 $H=(V_1,E_1)$，有 $V_1\subseteq V, E_1\subseteq E$，就称 H 是 G 的子图，记为 $H\subseteq G$。若 $H\subseteq G$ 且 $V(H)=V(G)$，就称 H 为 G 的生成子图，也可称 G 为 H 的生成母图。

定义 15 设图为图 G。对 G 的每一边 e，赋以一个实数 $w(e)$（如边的长度），称为 e 的权，由此得到赋权图 G。G 的子图的权是指子图的各边的权的和。

上述中国邮递员问题的数学模型是：在一个赋权连通图 G 上求一个含所有边的回路，且使此回路的权最小。

显然，若此连通赋权图 G 是 Euler 图，则可用 Fleury 算法求 G 的 Euler 回路，此回路即为所求的最优邮递员回路。

对于连通赋权图 G 为非 Euler 图的情形，1973 年，Edmonds 和 Johnson 给出了下面的解法，其实质是应用最短路算法，通过往图 G 中添加一些重复边，使图 G 成为 Euler 图 H，且使所添加的重复边的权之和最小，则 H 为 G 的最小权 Euler 生成母图。求得 H 的 Euler 回路，则此回路即为 G 的所求最优邮递员回路。其算法如下。

设 G 是连通赋权图，

(1)求 $V_0=\{v|v\in V(G), d(v)=1(\mathrm{mod}2)\}$。

(2)对每对顶点 $u,v\in V_0$，求 $d(u,v)$（$d(u,v)$是 u 与 v 的距离，可用 Floyd 算法求得）。

(3)构造完全赋权图 K_{V_0}，以 V_0 为顶点集，以 $d(u,v)$为边 uv 的权。

(4)求 $K_{|V_0|}$ 中权之和最小的完美对集 M。

(5)求 M 中边的端点之间的在 G 中的最短轨。

(6)在(5)中求得的每条最短轨上每条边添加一条等权的所谓“倍边”（即共端点共权的边）。

(7)在(6)中得到的图 G' 上求 Euler 回路即为中国邮递员问题的解。

注：若连通赋权图 G 恰好有两个奇点时，可以不用完美对集的内容。

例 8.6 一个求解中国邮递员问题的最小权 Euler 生成母图 H 的算例（见图 8-15）。

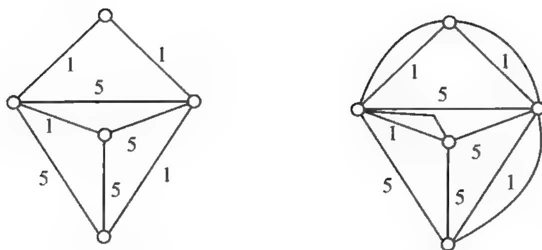


图 8-15 赋权图 G　　最小权 Euler 生成母图 H

以上算法的直观判别原理：对于连通赋权图 G 有多对奇点的情形可采用奇偶点图上作业法。

(1)找出图 G 中的所有奇顶点，把它们两两配成对，而每对奇点之间必有一条通路，把这条通路上的所有边作为重复边追加到图中去，这样得到的新连通图必无奇点。

(2)如果边 $e=(u,v)$上的重复边多于一条，则可从重复边中去掉偶数条，使得其重复边至多为一条，图中的顶点仍全部都是偶顶点。

(3)检查图中的每一个圈，如果每一个圈的重复边的总长不大于该圈总长的一半，则已经求得最优方案。如果存在一个圈，重复边的总长大于该圈总长的一半时，则将这个圈中的重复边去掉，再将该圈中原来没有重复边的各边加上重复边，其他各圈的边不变，返回步骤(2)。其原因是中国邮递员问题的解满足两条：在最优邮递路线上，图中的每一条边至多有一条重复边；在最优邮递路线上，图中每一个圈的重复边总权值小于或等于该圈总权值的一半。

8.4.3 多邮递员问题

邮局有 $k(k\geqslant2)$位投递员同时投递信件，全城街道都要投递，完成任务后返回邮局，如何分配投递路线，使得完成投递任务的时间最早？我们把这一问题记成 KPP。

KPP 的数学模型如下：

$G(V,E)$是连通图，$v_0\in V(G)$，求 G 的回路 $C_1,\cdots,C_k$，使得

(1)$v_0\in V(C_i)$，$i=1,2,\cdots,k$；

(2) $\max\limits_{1\leqslant i\leqslant k}\sum\limits_{e\in E(C_i)}w(e)=\min$；

(3) $\bigcup\limits_{i=1}^{k}E(C_i)=E(G)$。

8.4.4 中国邮递员问题的推广

我国南方雨量较多，部分地区经常遭受洪涝灾害，影响了人们正常的经济生产和日常生活。由此，出现了中国邮递员问题的推广问题——水灾地区邮递员问题，简称灾邮问题。

所谓灾邮问题，用非图论术语来讲，就是指在遭受水灾的地区，由于灾情的严重性，使得在某些道路上抗洪抢险部队只能保证通讯员安全通过一次，这就要求通讯员从总指挥部出发后，必须不重复地通过这些道路且遍历该地区的其他道路至少一次，最后以最短的路程完成投递任务返回总指挥部。

另外，中国邮递员问题还可应用到清洁车的城市清洁问题或扫雪车的城市扫雪问题中。

8.5 游山玩水问题

前面的 Euler 环游考虑的是经过图的每边恰好一次的回路，自然有人想到考虑经过图的每个顶点恰好一次的回路，这就是 Hamilton 圈问题，也有着广泛的应用。例如工序问题：今有工作 $J_1, J_2, \cdots, J_n$ 要在同一台机器上进行，从 J_i 到 J_j 机器调整时间为 t_{ij}，试把这 n 项工作排序，使得 $\sum t_{ij}$ 最小。又如旅行商(TSP)问题(也称货郎问题)：一名推销员准备前往若干城市推销产品，然后回到他的出发地。如何为他设计一条最短的旅行路线(从驻地出发，经过每个城市恰好一次，最后返回驻地)？这些问题用图论的术语说，就是在一个赋权完全图中，找出一个有最小权的 Hamilton 圈(称这种圈为最优圈)。

8.5.1 Hamilton 圈和 Hamilton 图

包含 G 的每个顶点的轨叫做 Hamilton(哈密顿)轨；闭的 Hamilton 轨叫做 Hamilton 圈或 H 圈；含 Hamilton 圈的图叫做 Hamilton 图。

直观地讲，Hamilton 图就是从一顶点出发、每顶点恰通过一次能回到出发点的那种图，即不重复地行遍所有的顶点再回到出发点。

与上节的问题不同，目前还没有求解旅行商问题的有效算法。所以希望有一个方法以获得相当好(但不一定最优)的近似解。但当赋权完全图的权满足三角不等式等一些条件时，有一些比较好的近似算法，可参考其他文献。

一个可行的办法是首先用最近邻域算法(NN)求一个 Hamilton 圈 C，然后适当修改 C 以得到具有较小权的另一个 Hamilton 圈。修改的方法叫做改良圈算法。

8.5.2 求 Hamilton 圈和 Hamilton 图的近似算法

第一种，最近邻域算法(NN)。

(1) 任选一个城市作为部分旅游(v_1)。

(2) 设部分旅游为($v_1v_2\cdots v_i$)，若 v_{i+1} 是不在部分旅游($v_1v_2\cdots v_i$)中的所有城市里离 v_i 最近的城市，则将它加入部分旅游($v_1v_2\cdots v_i$)中，得新的部分旅游($v_1v_2\cdots v_iv_{i+1}$)。

(3) 如果当前的部分旅游已经含有所有的旅游城市，则停止，否则转(2)，从而得初始圈 $C=v_1v_2\cdots v_nv_1$。

第二种，改良圈算法。

设初始圈 $C=v_1v_2\cdots v_nv_1$，对于 $1<i+1<j<n$，构造新的 Hamilton 圈：

$C_{ij}=v_1v_2\cdots v_iv_jv_{j-1}v_{j-2}\cdots v_{i+1}v_{j+1}v_{j+2}\cdots v_nv_1$，它是由 C 中删去边 v_iv_{i+1} 和 v_jv_{j+1}、添加边 v_iv_j 和 $v_{i+1}v_{j+1}$ 而得到的。若 $w(v_iv_j)+w(v_{i+1}v_{j+1})<w(v_iv_{i+1})+w(v_jv_{j+1})$，则以 C_{ij} 代替 C，C_{ij} 叫做 C 的改良圈。

循环运算，直至无法改进，停止。

用改良圈算法得到的结果几乎可以肯定不是最优的。为了得到更高的精确度，可以选择不同的初始圈，重复进行几次算法，以求得较精确的结果。

这个算法的优劣程度有时能用 Kruskal 算法加以说明。假设 C 是 G 中的最优圈，则对于任何顶点 v，$C-v$ 是在 $G-v$ 中的 Hamilton 轨，因而也是 $G-v$ 的生成树。由此推知：若 T 是 $G-v$ 中的最优树，同时 e 和 f 是和 v 关联的两条边，并使得 $w(e)+w(f)$ 尽可能小，则 $w(T)+w(e)+w(f)$ 将是 $w(C)$ 的一个下界。

这里介绍的方法已被进一步发展。在圈的修改过程中一次替换三条边比一次仅替换两条边更为有效，然而，有点奇怪的是，进一步推广这一想法，就不利了。

例 8.7 从北京(Pe)乘飞机到东京(T)、纽约(N)、墨西哥城(M)、伦敦(L)、巴黎(Pa)五个城市旅游，每城市恰去一次再回北京，应如何安排旅游线，使旅程最短？各城市之间的航线距离如下表：

表 8-2 六城市间航线距离

	L	M	N	Pa	Pe	T
L		56	35	21	51	60
M	56		21	57	78	70
N	35	21		36	68	68

续表

Pa	21	57	36		51	61
Pe	51	78	68	51		13
T	60	70	68	61	13	

解：编写程序如下：

```
clc,clear
a(1,2)=56;a(1,3)=35;a(1,4)=21;a(1,5)=51;a(1,6)=60;
a(2,3)=21;a(2,4)=57;a(2,5)=78;a(2,6)=70;
a(3,4)=36;a(3,5)=68;a(3,6)=68;
a(4,5)=51;a(4,6)=61;
a(5,6)=13;
a(6,:)=0;
a=a+a';
c1=[5 1:4 6];
L=length(c1);
flag=1;
while flag>0
      flag=0;
   for m=1:L-3
      for n=m+2:L-1
         if a(c1(m),c1(n))+a(c1(m+1),c1(n+1))<a(c1(m),c1(m+1))+a(c1
(n),c1(n+1))
            flag=1;
            c1(m+1:n)=c1(n:-1:m+1);
         end
      end
   end
end
sum1=0;
for i=1:L-1
   sum1=sum1+a(c1(i),c1(i+1));
end
circle=c1;
sum=sum1;
c1=[5 6 1:4];%改变初始圈，该算法的最后一个顶点不动
flag=1;
```

```
while flag>0
      flag=0;
  for m=1:L-3
    for n=m+2:L-1
      if a(c1(m),c1(n))+a(c1(m+1),c1(n+1))<...
          a(c1(m),c1(m+1))+a(c1(n),c1(n+1))
        flag=1;
        c1(m+1:n)=c1(n:-1:m+1);
      end
    end
  end
end
sum1=0;
for i=1:L-1
  sum1=sum1+a(c1(i),c1(i+1));
end
if sum1<sum
  sum=sum1;
  circle=c1;
end
circle,sum
```

可以估计得其旅程的下界为 122+21+35=178。

8.5.3 旅行商(TSP)问题的推广

从服务型的角度推广旅行商(TSP)问题，可产生多旅行商问题：多名服务员准备前往若干个城市进行产品服务，然后回到他的出发地。如何为他们设计相应旅行路线最短的行程方案，请参考其他有关文献。

思考题

1. 求图 8-7 的欧拉道路。

2. 请你为图 8-16 所示街区的邮递员设计一个送信方案(邮局在 A 点处)。

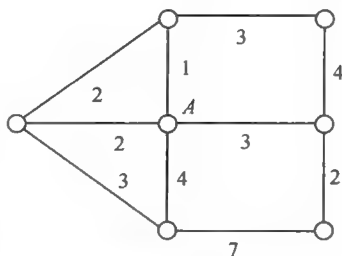


图 8-16

3. “5·12”汶川大地震使震区地面交通和通讯系统严重瘫痪。救灾指挥部紧急派出多支小分队，到各个指定区域执行搜索任务，以确定需要救助的人员的准确位置。在其他场合也常

有类似的搜索任务。在这种紧急情况下需要解决的重要问题之一是：制定搜索队伍的行进路线，对预定区域进行快速的全面搜索。通常，每个搜索人员都带有GPS定位仪、步话机以及食物和生活用品等装备。队伍中还备有一定数量的卫星电话。GPS可以让搜索人员知道自己的方位。步话机可以相互联系。卫星电话用来向指挥部报告搜索情况。

下面是一个简化的搜索问题。有一个平地矩形目标区域，大小为12000 m×7000 m，需要进行全境搜索。假设：出发点在区域中心；搜索完成后需要进行集结，集结点（结束点）在左侧短边中点；每个人搜索时的可探测半径为20 m，搜索时平均行进速度为0.6 m/s；不需搜索而只是行进时，平均速度为1.2 m/s。每个人带有GPS定位仪、步话机，步话机通讯半径为1000 m。搜索队伍若干人为一组，有一个组长，组长还拥有卫星电话。每个人搜索到目标，需要用步话机及时向组长报告，组长用卫星电话向指挥部报告搜索的最新结果。

现在有如下问题需要解决。

(1)假定有一支20人一组的搜索队伍，拥有1台卫星电话。请设计一种你认为耗时最短的搜索方式。按照你的方式，搜索完整个区域的时间是多少？能否在48小时内完成搜索任务？如果不能完成，需要增加到多少人才可以完成？

(2)为了加快速度，搜索队伍有50人，拥有3台卫星电话，分成3组进行搜索。每组可独立将搜索情况报告给指挥部门。请设计一种你认为耗时最短的搜索方式。按照你的搜索方式，搜索完整个区域的时间是多少？

4. 图8-17为某县的乡（镇）、村公路网示意图，公路边的数字为该路段的公里数。某年夏天该县遭受水灾。为考察灾情、组织自救，县领导决定带领有关部门负责人到全县各乡（镇）、村巡视。巡视路线指从县政府所在地出发，走遍各乡（镇）、村，又回到县政府所在地的路线。

(1)若分三组（路）巡视，试设计总路程最短且各组尽可能均衡的巡视路线。

(2)假定巡视人员在各乡（镇）停留时间$T=1.5$ h，在各村停留时间$t=0.45$ h，汽车行驶速度$V=30$ km/h。要在24 h内完成巡视，至少应分几组？给出这种分组下你认为最佳的巡视路线。

(3)在上述关于T，t和V的假定下，如果巡视人员足够多，完成巡视的最短时间是多少？给出在这种最短时间完成巡视的要求下，你认为最佳的巡视路线。

(4)若巡视组数已定（比如3组），要求尽快完成巡视，讨论T，t和V改变

对最佳巡视路线的影响。

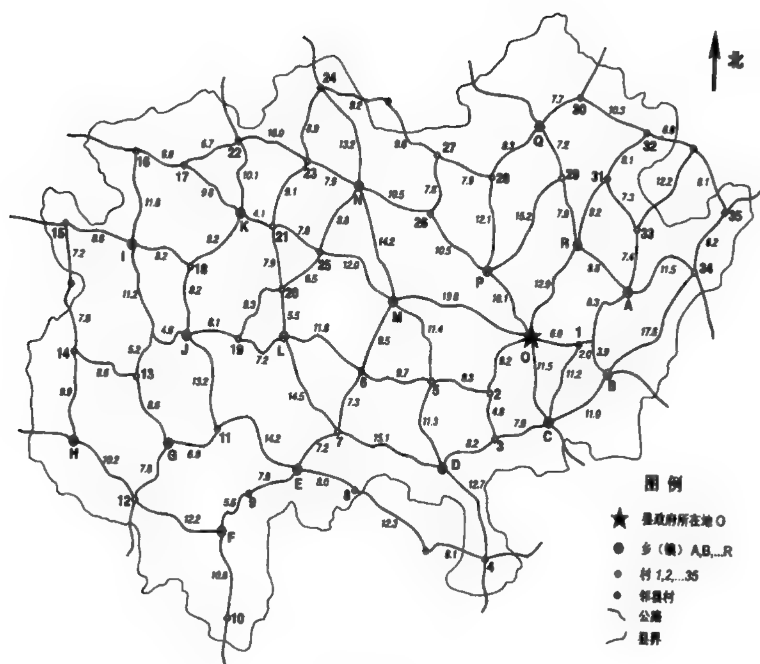


图 8-17

第9章 密码学及其应用

密码(cipher, cryptogram, secret code)是一种用来混淆的技术,它希望将正常的(可识别的)信息转变为无法识别的信息。当然,对一小部分人来说,这种无法识别的信息是可以再加工并恢复的。"密码"一词对人们来说并不陌生,人们可以举出许多有关使用密码的例子。如保密通信设备中使用"密码",个人在银行取款使用"密码",在计算机登录和屏幕保护中使用"密码",开启保险箱使用"密码",电子游戏中使用"密码"等。"密码"在中文里是"口令(password)"的通称。登录网站、电子邮箱和银行取款时输入的"密码"其实严格来讲应该被称作"口令",因为它不是本来意义上的"加密代码",但是也可以称为秘密的号码。

密码在古代就被用于传递秘密消息;在近代和现代战争中,传递情报和指挥战争均离不开密码,外交斗争中也离不开密码。密码一般用于信息通信传输过程中的保密和存储中的保密。随着计算机和信息技术的发展,密码技术的发展也非常迅速,应用领域不断扩展。密码除了用于信息加密外,也用于数据信息签名和安全认证。这样,密码的应用也不再只局限于为军事、外交斗争服务,它也广泛应用在社会和经济活动中。当今世界已经出现了密码应用的社会化和个人化趋势。例如,可以将密码技术应用在电子商务中,对网上交易双方的身份和商业信用进行识别,防止网上电子商务中的"黑客"和欺诈行为;应用于增值税发票中,可以防伪、防篡改,杜绝了各种利用增值税发票偷、漏、逃、骗国家税收的行为,并大大方便了税务稽查;应用于银行支票鉴别中,可以大大降低利用假支票进行金融诈骗的金融犯罪行为;应用于个人移动通信中,大大增强了通信信息的保密性,等等。

9.1 密码故事

9.1.1 雅典密码

公元前405年,雅典和斯巴达之间的伯罗奔尼撒战争已进入尾声。斯巴达军队逐渐占据了优势地位,准备对雅典发动最后一击。这时,原来站在斯

巴达一边的波斯帝国突然改变态度，停止了对斯巴达的援助，意图是使雅典和斯巴达在持续的战争中两败俱伤，以便从中渔利。在这种情况下，斯巴达急需摸清波斯帝国的具体行动计划，以便采取新的战略方针。正在这时，斯巴达军队捕获了一名从波斯帝国回雅典送信的雅典信使。斯巴达士兵仔细搜查这名信使，可搜查了好大一阵，除了从他身上搜出一条布满杂乱无章的希腊字母的普通腰带外，别无他获。情报究竟藏在什么地方呢？斯巴达军队统帅莱桑德把注意力集中到了那条腰带上，情报一定就在那些杂乱的字母之中。他反复琢磨研究这些天书似的文字，把腰带上的字母用各种方法重新排列组合，怎么也解不出来。最后，莱桑德失去了信心，他一边摆弄着那条腰带，一边思考着弄到情报的其他途径。当他无意中把腰带呈螺旋形缠绕在手中的剑鞘上时，奇迹出现了。原来腰带上那些杂乱无章的字母，竟组成了一段文字。这便是雅典间谍送回的一份情报，它告诉雅典，波斯军队准备在斯巴达军队发起最后攻击时，突然对斯巴达军队进行袭击。斯巴达军队根据这份情报马上改变了作战计划，先以迅雷不及掩耳之势攻击毫无防备的波斯军队，并一举将它击溃，解除了后顾之忧。随后，斯巴达军队回师征伐雅典，取得了战争的最后胜利。

雅典间谍送回的腰带情报，被认为是世界上最早的密码情报，具体运用方法是，通信双方首先约定密码解读规则，然后通信一方将腰带（或羊皮等其他东西）缠绕在约定长度和粗细的木棍上书写。收信一方接到后，如不把腰带缠绕在同样长度和粗细的木棍上，就只能看到一些毫无规则的字母。后来，这种密码通信方式在希腊广为流传。现代的密码电报，据说就是受了它的启发而发明的。

9.1.2 圣经密码

圣经密码，也称作 Torah 密码，最初指的是在旧约《创世记》的开头每隔 50 个字母跳读，就可以拼出“Torah”一词（意指摩西五经，《创世记》《出埃及记》《利未记》《民数记》及《申命记》），另外在《出埃及记》《民数记》和《申命记》中亦是如此。这种现象后来被称作等距字母序列（Equidistant Letter Sequences，简称“ELS”）。这个密码，由于 *The Bible Code* 一书的出版而闻名于世，书中作者声称这些密码可以预言将来。这被无神论者和许多宗教团体强烈质疑。

后来由以色列数学界被认为天才的 Eliyahu Rips 和物理学家 Doron Witztum 利用电脑，从圣经时代到现代的智者中，挑出 32 位大圣贤的名字和他们出生与死亡日期，发现在《创世纪》里面都是编在一起的。但其他书本却

不能找到这个规律。

其实,大科学家牛顿在生命最后一段的工作中,已相信《圣经》暗藏着人类历史的预言,但可惜他毕生的探求终归徒然。

9.2 密码学发展简史

从整体来说,密码学经历了从人工密码到机械密码再到电子计算机的发展历程。

9.2.1 古代加密密码——手工阶段

(1)公元前 440 年,古希腊战争中出现了隐写术。当时为了安全传送军事情报,奴隶主剃光奴隶的头发,将情报写在奴隶的光头上,待头发长长后将奴隶送到另一个部落,再次剃光头发,原有的信息浮现出来,从而实现了两个部落间的秘密通信。

(2)公元前 1 世纪,古罗马帝国皇帝恺撒曾经使用有序的单表代替密码,即恺撒密码。

(3)在我国唐宋时代,也有以藏头诗、藏尾诗、漏格诗及绘画等形式,将要表达的真正意思或“密码”隐藏在诗文或画卷中特定位置的记载,这种方式已具备了密码本的特点。

9.2.2 古典密码——机械阶段

古典密码的加密方法,一般是文字置换,使用手工或机械变换的方式实现。它比古代加密方法复杂,其变化较小,古典密码的代表密码体制主要有:单表代替密码、多表代替密码和转轮密码。

(1)阿拉伯人是第一个清晰地理解密码学原理的人,他们设计并使用代换和换位加密,并且发现了密码分析中的字母频率分布关系。

(2)欧洲的密码学起源于中世纪的罗马和意大利,约在 1379 年,欧洲第一本关于密码学的手册由 Gabriela de Lavinde 编写。由几个加密算法组成,并且为罗马教皇 Clement 七世服务。

(3)到了 1860 年,密码系统在外交通信中已得到普遍使用。如在美国国内战争期间,联邦军中广泛使用了换位加密,主要使用的是 Vigenere 密码,并且偶尔使用单字母代替,然而联合军密码分析人员破译了截获的大部分联邦军密码。

在第一次世界大战期间,双方都使用加密系统。主要用于战术通信,一

些复杂的加密系统被用于高级通信中，直到战争结束。而密码本系统，主要用于高级命令和外交通信中。

(4)20 世纪 20 年代，随着机械和机电技术的成熟以及电报和无线电需求的出现，引起了密码设备方面的一场革命——发明了转轮密码机(Rotor)。转轮密码机的出现是密码学发展的重要标志之一，从而出现了商业密码机的公司和市场。

转轮机由一个键盘和一系列转轮组成，每个转轮是 26 个字母的任意组合。转轮被齿轮连接起来。当一个转轮转动时，可以将一个字母转换为另一个字母，依此传递下去，当最后一个转轮处理完成时，就可以得到加密后的字母。为了使转轮密码更安全，人们还把几种转轮和移动齿轮结合起来，所有转轮以不同的速度转动，并且通过调整转轮上字母的位置和速度为破译设置更大的障碍。

(5)第二次世界大战是人工加密时代转为机械加密时代的转折点，转轮密码机的大量使用极大地提高了加密解密的速度。抗攻击性能的极大提高是密码学发展史上的一个里程碑。二次大战期间是密码分析获得最大成功的时期。波兰人和英国人破译了德国的 Enigma 密码，美国人攻破了日本的 RED、ORANGE 和 PURPLE 密码，对盟军在二次大战中获胜起到了关键性作用。

9.2.3 现代密码——计算机阶段

(1)1949 年 Shannon 发表了《保密系统的通信理论》(*Communication Theory of Secrecy System*)一文，用概率和统计等科学工具研究加密系统，为密钥学奠定了坚实的理论基础。从而密码学从艺术变为科学。当然，此文发表了 30 年后才显示出它的价值。

(2)1976 年，W. Dffie 和 M. Hellman 发表了《密码学的新方向》，提出了公钥密码体制，这是密码学发展史上最伟大的一次革命，是现代密码学诞生的标志。

(3)1978 年，美国麻省理工学院(MIT)的 Rivest、Shamir 和 Adleman 提出 RSA 加密机制，这是第一个实用的公钥加密方案，开创了密码学的新纪元。

现代密码学与计算机技术、电子通信技术紧密相关。在这一阶段，密码理论蓬勃发展，密码算法设计与分析互相促进，出现了大量的密码算法和各种攻击方法。另外，密码使用的范围也在不断扩展，而且出现了一些新的密码技术，如混沌密码、量子密码等。

9.3 密码学基本知识

密码的设计和使用至少可以追溯到4000多年前的埃及、巴比伦、罗马和希腊,历史极为悠久。古代隐藏信息的方法主要有两大类:一类为隐藏信息载体,采用隐写术等;另一类为变换信息载体,使之无法为一般人所理解。本节只涉及后者,介绍一些采用比较简单的数学工具对信息加密、解密的技巧和方法。

在密码学中,信息代码被称为密码,加密前的信息称为明文(plaintext),经加密后的信息称为密文(ciphertext),一般是不为人所理解的信息。将明文转变为密文的过程称为加密(enciphering),其逆过程称为解密(deciphering),而用以加密、解密的方法或算法则称为密码体制(crytosystem)。

全体明文组成的集合称为明文空间,记作U,全体密文组成的集合称为密文空间,记作V。加密常采用密钥来实现,密钥通常取自于一个被称为密钥空间的含有若干参数的集合K。按数学的观点来看,加密与解密均是一种变换:取一个$k\in K$,$\forall u\in U$,令$u\xrightarrow{k}v\in V$,v是明文u在密钥K下的密文;而解密为加密K的逆变换K^{-1},$v\xrightarrow{k^{-1}}u$。由此可见,尽管密码体系可能千姿百态,但其关键还在于密钥的选取。

随着计算机与网络技术的迅猛发展,大量各具特色的密码体系不断涌现。离散数学、数论、计算复杂性、椭圆曲线、混沌与分形等理论都已用到现代密码学理论中。本文只对古典密码、希尔密码和公钥密码作简单介绍,前两者用到了简单的线性代数运算,后者用到了简单的数论知识。

早期密码大体分为替代密码和代数密码。先简单介绍替代密码,然后再介绍一类重要的代数密码——希尔密码。

9.4 代替密码

9.4.1 替代密码

替代密码也称为置换密码,是最简单的加密和解密方法,采用另一个字母表中的字母代替明文中的字母,明文字母与密文字母保持一一对应关系,但采用的符号改变了。加密时,把明文用密文代替,即把明文中的字母用密文字母表中对应的字母代替;解密则与此相反,通过对应关系,将密文换回成明文。解密与加密是互逆过程。在替代法加密过程中,密文字母表称为密钥,可以是标准字母表,也可以是按任意顺序建立的字母表。以英文为例,英

文字母共有 26 个字母，在不重复的前提下，每一个字母都可以用另一个字母替代，所以，共可构造出 26！种不同的字母置换表，加密和解密时必须选定其中的同一张置换表。由于有密钥（即置换字母表），加密和解密都比较容易实现，但对于不知道这张置换表的破译者来说绝非易事，因为他只能逐一尝试全部张可能的置换字母表。

(1)单表代替密码

单表代替密码对明文中的所有字母都使用同一个映射，即

$$\forall p \in P,\ f: P \rightarrow C,\ f(p)=c$$

为了保证加密的可逆性，一般要求映射 f 为一一映射。单表代替包括最早的恺撒加密（加密时向后移 3 位，由于 3 是固定的，故没有密钥）和一般意义上的单字母代替（即移位密码，也称通用恺撒密码，向后移 k 位，k 是任意的，故认为 k 是密钥，共有 26 个密钥）及使用密钥的单表代替与仿射加密。

①移位密码

最早的代替密码是移位密码，现在为人所知的最为古老的加密方法——天书，即是移位法的一种。早在 4000 多年前，古希腊人用一种名叫“天书”的器械来加密消息。该密码是用一条窄长的草纸缠绕在一个直径确定的圆筒上，明文逐行横写在纸带上，当取下纸带时，字母的次序被打乱了，消息得以屏蔽。收方阅读消息时，要将纸带重新绕在直径与原来相同的圆筒上，才能看到正确的消息。当然，这里的圆筒起到了密钥的作用。

另一种移位法采用将字母表中的字母平移若干位的方法来构造密文字母表，传说这类方法是由古罗马执政官恺撒大帝最早使用的，所以这种密文字母表也称为恺撒字母表，相应的密码也称为恺撒密码。其转换方式为：

$$q=p+c(\text{mod } 26)$$

其中，c 为事先取定的常数，也就是恺撒密码的密钥；p 为明文字母；q 为 p 对应的密文字母（对英文加密，并以 26 为模取同余运算）。

例 9.1 取 $c=10$，可得以下的字母对应表：

明文字母表	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
密文字母表	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J

利用此表，可将明文“Good morning”加密为密文“Qyyn wybxsxq”。

此法加密简单，解密也简单，只要作加密运算的逆运算就可以，即令

$$p \equiv q-c(\text{mod}26)$$

就可以很容易将密文“Qyyn wybxsxq”解密为“Good morning”。

正因为恺撒密码加密、解密都很简单，所以敌人破译也较为容易。一旦

破译者知道使用者使用的是恺撒密码，则只需知道密钥 c 就可以破译。当然，破译者也可以令 $c=1, 2, \cdots, m-1$，一一加以试验，直到获取明文。

恺撒密码的特点：仅对字母表进行了平移，保留了原字母表中的相邻关系，也为破译提供了信息。

②使用密钥的单表代替加密

这种密码选用一个英文短语或单词串作为密钥，去掉其中重复的字母得到一个无重复字母的字母串，然后再将字母表中的其他字母依次写于此字母串之后，就可以构造一个字母代替表。这种单表代替泄漏破译者的信息更少，而且密钥可以随时更改，增加了灵活性。

例 9.2 设密钥为：construct

明文字母表	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
密文字母表	c	o	n	s	t	r	u	a	b	d	e	f	g	h	i	j	k	l	m	p	q	v	w	x	y	z

因此，如果明文为“China”，则对应的密文为“nabhc”。

在设计密钥时，也可在明文字母表中选择一个特定字母，然后从该特定字母开始写密钥字将密钥字隐藏于其中。

例如，对于上例，选取特定字母 k，则可得

明文字母表	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
密文字母表	k	l	m	p	q	v	w	x	y	z	c	o	n	s	t	r	u	a	b	d	e	f	g	h	i	j

因此，如果明文为“China”，则对应的密文为“mxysk”。

③仿射加密

为了增强保密性，有人提出了仿射变换密码，公式为

$$\text{加密}: q \equiv ap + b \pmod{26},$$

$$\text{解密}: p \equiv a^{-1}(q-b) \pmod{26},$$

其中，a^{-1} 为 a 的逆元素。其中，a，b 为两个事先选定的常数。所以 a 有 12 个取值：1,3,5,7,9,11,15,17,19,21,23,25；b 有 26 个取值。

显然，当 $a=1$ 时为通用恺撒变换，如果同时 $b=3$，则为恺撒变换。

仿射加密要求 $\gcd(a, 26)=1$，即 a 与 26 互素，否则就退化为 $q \equiv b \pmod{26}$。

例 9.3 取 $a=3$，$b=0$，例 9.1 中的“Good morning”可以加密为“Ussl msbpapu”。

为了计算方便，先用 0～25 代替字母 A～Z，得到以下对应表：

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	0
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

在 $a=3$，$b=0$ 时，置换字母表为：

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
C	F	I	L	O	R	U	X	A	D	G	J	M	P	S	V	Y	B	E	H	K	N	Q	T	W	Z

从字母置换表可以看出，仿射变换与移位变换相比，已有了较大的改进，在仿射变换的字母表中，字母间原有的顺序被改变，从而增加了破译难度。

例 9.4 设$(a, b)=(7, 3)$，注意到$7^{-1}(\bmod 26)=15$，加密函数$q=7p+3(\bmod 26)$，相应的解密函数为 $p=15(q-3)(\bmod 26)=15q-19(\bmod 26)$。

若要加密明文“China”，首先，按照字母表中顺序，从 0 开始编号，直到 26，将字母 C,h,i,n,a 转换成 2,7,8,13,0，然后加密：

$$7\times\begin{bmatrix}2\\7\\8\\13\\0\end{bmatrix}+\begin{bmatrix}3\\3\\3\\3\\3\end{bmatrix}=\begin{bmatrix}17\\52\\59\\94\\3\end{bmatrix}\bmod 26=\begin{bmatrix}17\\0\\7\\16\\3\end{bmatrix}-\begin{bmatrix}R\\A\\H\\Q\\D\end{bmatrix}$$

即在当前密钥下，“China”经仿射加密变换为“RAHQD”。

解密：

$$15\times\begin{bmatrix}17\\0\\7\\16\\3\end{bmatrix}-\begin{bmatrix}19\\19\\19\\19\\19\end{bmatrix}=\begin{bmatrix}236\\-19\\86\\221\\26\end{bmatrix}\bmod 26=\begin{bmatrix}2\\7\\8\\13\\0\end{bmatrix}=\begin{bmatrix}C\\H\\I\\N\\A\end{bmatrix}$$

原始信息“China”已得到恢复。

以上介绍的恺撒密码和仿射变换密码均属于置换密码，置换密码有一个共同的缺点，即将明文字母转换为密文字母表时，字母间的对应关系是一对一的，这就使得明文中的某些特征(例如字母的频率特征、语法特征等)在密文中保存了下来，为破译密文提供了很有参考价值的线索。结合这些特征，再使用穷举法，最终可以获取明文。所以，将穷举法与统计法结合起来分析是破译代替法密码所采用的最基本和常用的方法，其他特殊方法大多是这两种方法的综合与改进。

(2)多表代替密码

单表代替密码表现出明文中单字母出现频率分布与密文中相同，多表代替密码使用从明文字母到密文字母的多个映射来隐藏单字母出现的频率分布，每个映射是简单代替密码中的一对一映射(即处理明文消息时使用不同的单字母代替)。多表代替密码将明文字符划分为长度相同的消息单元，称为明文组，对字符块成组进行代替，即使用了多张字符代替表，这样一来同一个字符具有不同的密文，改变了单表代替中密文的唯一性，使密码分析更加困难。多字母代替的优点是很容易将字母的自然频度隐藏或均匀化，从而有利于抗击统计分析。

在多字母代替中，每一组有 D 个字母，每个字母有 26 种可能，所以秘密取决于 D, f_1, f_2, …, f_D。对于每一个变换 f，有 26! 种可能，共有 D 个变换。所以密钥总数为 $26!\times 26!\times\cdots\times 26!=(26!)^D$。Playfair 密码、Hill 密码、Vigenere 密码都是多表代替密码，本文只介绍 Hill(希尔)密码。

希尔密码(Hill 密码)是由数学家 Lester Hill 于 1929 年发明的。算法的基本思想：将 D 个连续的明文字母通过线性变换转换为 D 个密文字母。这种代替由 D 个线性方程决定，其中每个字母被分配一个数值(0, 1, …, 25)。解密只需做一次逆变换即可。密钥就是变换矩阵本身。即：

$$\text{明文}:M=M_1M_2\cdots M_d$$

$$\text{密文}:C=E(M)=C_1C_2\cdots C_d$$

其中，$C_1=k_{11}M_1+k_{12}M_2+\cdots+k_{1d}M_d$

$C_2=k_{21}M_1+k_{22}M_2+\cdots+k_{2d}M_d$

……

$C_d=k_{d1}M_1+k_{d2}M_2+\cdots+k_{dd}M_d$

写成矩阵形式为：

$$C_{1,d}=M_{1,d}\cdot K_{d,d}$$

或 $$(c_1,c_2,\cdots,c_d)=(m_1,m_2,\cdots,m_d)\cdot\begin{pmatrix}k_{11} & k_{12} & \cdots & k_{1d}\\ \vdots & \vdots & & \vdots\\ k_{d1} & k_{d2} & \cdots & k_{dd}\end{pmatrix}$$

即密文分组＝明文分组×密钥矩阵。

例 9.5 明文为：$P=\text{HILL}$，使用的密钥为：

$$K=\begin{pmatrix}8 & 6 & 9 & 5\\ 6 & 9 & 5 & 10\\ 5 & 8 & 4 & 9\\ 10 & 6 & 11 & 4\end{pmatrix}$$

HILL 被数字化后的 4 个数字为:7,8,11,11(字母表中顺序,编号从 0 开始)。

$$所以,C=(7\quad 8\quad 11\quad 11)\cdot\begin{pmatrix}8&6&9&5\\6&9&5&10\\5&8&4&9\\10&6&11&4\end{pmatrix}\bmod 26$$

$$=(9,8,8,24)=(\text{JIIY})$$

由线性代数方法,可以得到矩阵 K 的逆矩阵:

$$K^{-1}=\begin{pmatrix}23&20&5&1\\2&11&18&1\\2&20&6&25\\25&2&22&25\end{pmatrix}$$

所以解密有:

$$P=CK^{-1}=(9\quad 8\quad 8\quad 24)\cdot\begin{pmatrix}23&20&5&1\\2&11&18&1\\2&20&6&25\\25&2&22&25\end{pmatrix}\bmod 26$$

$$=(7,8,11,11)=(\text{HILL})$$

显然,将长消息分组,分组的长度由矩阵的维数决定。希尔密码的强度在于完全隐藏了单字母的频率。字母和数字的对应也可以改成其他方案,使其攻击更不易成功。一般来说,希尔密码能比较好地抵抗频率法分析,对抗仅有密文的攻击强度较高,但易受已知明文攻击。

9.5 RSA 密码体制

9.5.1 RSA 密码体制基本概念及基本定理

在有些情况下,事先约定好密钥是不可能的,例如某个工程项目发出招标公告,希望有尽可能多的公司参与投标,并使用通讯方式进行洽谈。此时既需要保守商业秘密(对第三方),又不可能事先约定所使用的密钥。公开密钥体制的提出就是为了从根本上解决上述问题。其基本思想是:把密钥划分为公开密钥和秘密密钥两部分,两者互为逆变换,但几乎不可能从公开密钥推出秘密密钥。每个使用者均有自己的公开密钥和秘密密钥。公开密钥供别人向自己发送信息时加密使用,这种密钥可以像电话号码一样提供给一切人查阅;而每个用户对自己的秘密密钥则严加保密,只供自己解密使用。这其中就要利用某些已知的 NPC 问题,例如背包问题、整数分解问题、矩阵覆盖

问题等。

RSA 体制是一种当前还在使用的公开密钥方法,用三个发明者里弗斯特(Rivest)、沙米尔(Shamir)和艾德曼(Adleman)的名字首字母命名。这一方法的安全性基于如下简单的事实:将两个大素数相乘很容易,但反过来将它们的乘积再因式分解回来则相当困难。

定义 1 设 n 为一正整数,将小于 n 且与 n 互素的正整数个数记为$\varphi(n)$,称此函数为欧拉(Euler)φ 函数。

显然,$\varphi(2)=1$,因为 1 与 2 互素;$\varphi(3)=2$,因为 1, 2 都与 3 互素;$\varphi(4)=2$,因为 1, 3 都与 4 互素。不难发现:若 p, q 为两个相异素数,$n=pq$,则 $\varphi(n)=(p-1)(q-1)$。

令 p, q 为随机选取的两个大素数(大约为十进制 100 位或更大),令 $n=pq$,n 是公开的,而 p, q 是保密的。掌握 $n=pq$ 的人知道欧拉函数值 $\varphi(n)=(p-1)(q-1)$,但若不知道因式分解就不能用这个公式计算。随机选取一个数 e,e 为小于 $\varphi(n)$且与它互素的正整数。利用辗转相除法,可以找到整数 d 和 r,使得

$$ed+r\varphi(n)=1,\text{即 } ed\equiv 1(\bmod\ \varphi(n))$$

数 n,e 和 d 分别称为模、加密密钥和解密密钥。数 n 和 e 组成公开密钥的加密密钥,而其余的项 p, q, $\varphi(n)$和 d 组成了秘密陷门。显然,陷门信息包含了四个相关的项,如果知道了 p,其余三项立刻可以求出。如果不知道 $\varphi(n)$,由已知的公开密钥 e,很难算出秘密密钥 d。若知道了 $\varphi(n)$,则由

$$pq=n,\ p+q=n-\varphi(n)+1$$

可知 p, q 是二次方程 $x^2+[\varphi(n)-n-1]x+n=0$ 的根,由此可以求出 p, q。从而将 n 因式分解。所以,RSA 体制的安全性与因式分解密切相关,若能知道 n 的因式分解,该密码就能破译。因此,要选用足够大的 n,使得在当今的条件下要分解它是十分困难的。

RSA 体制常用的定理有:

定理 1 (欧拉定理)若 $\gcd(a, n)=1$,则 $a^{\varphi(n)}\equiv 1(\bmod\ n)$。

定理 2 (费马小定理)若 p 是素数,则对一切整数 a 有 $a^p\equiv a(\bmod\ p)$。

9.5.2 RSA 方案描述

(1)密钥生成

①选取两个大素数 p 和 q(例如长度都接近 512 bit)。

②计算乘积 $n=p\times q$, $\varphi(n)=(p-1)(q-1)$,其中 $\varphi(n)$为 n 的欧拉函数。

③随机选择整数 $e(1<e<\varphi(n))$,要求满足 $\gcd(e, \varphi(n))=1$,即 e 与$\varphi(n)$

互素。

④用扩展的 Euclidean 算法(辗转相除法)计算私钥 d,以满足 $d\times e\equiv 1 \bmod (\varphi(n))$,即 $d\equiv e^{-1}\bmod(\varphi(n))$。

得到:公钥为 e 和 n,d 是私钥(两个素数 p 和 q,可销毁,不能泄露)。

(2)加密过程

明文先转换为比特串分组,使每个分组对应的十进制数小于 n,即分成长度小于 $\log_2 n$,然后对每个明文分组 m_i 作加密运算,具体过程如下。

①获得接收公钥(e, n)。

②把消息 M 分成长度为 $L(L<\log_2 n)$的消息分组 $M=m_1m_2\cdots m_t$。

③使用加密算法 $c_i=m_i^e \bmod n\ (1\leqslant i\leqslant t)$,计算出密文 $c=c_1c_2\cdots c_t$。

(3)解密过程

①将密文 c 按长度 L 分组得 $c=c_1c_2\cdots c_t$。

②使用私钥 d 和解密算法$m_i=c_i^d \bmod n\ (1\leqslant i\leqslant t)$计算 m_i。

③得明文消息 $M=m_1m_2\cdots m_t$。

例 9.6 取 $p=11$,$q=13$,则 $n=pq=11\times 13=143$,$\varphi(n)=(p-1)(q-1)=120$,选取 $e=17$,满足 $\gcd(e, \varphi(n))=\gcd(17, 120)=1$。使用扩展的 Euclidean 算法计算,$d=e^{-1}=113 \bmod 120$,所以公钥为$(n, e)=(143, 17)$,私钥为 $d=113$。设对明文 $m=24$ 进行加密,密文为 $c\equiv m^e\equiv 24^{17}\equiv 7 \bmod 143$。密文 $c=7$ 经公开信道发送到接收方后,接收方用私钥 $d=113$ 对密文解密:$m\equiv c^d\equiv 7^{113}\equiv 24 \bmod 143$。从而恢复明文。

例 9.7 取 $p=47$,$q=59$,则 $n=pq=47\times 59=2773$,$\varphi(n)=(p-1)(q-1)=2668$,选取 $e=17$,满足 $\gcd(e, \varphi(n))=\gcd(17, 2668)=1$,且易得到 $d=157$。将$(n, e)=(2773, 17)$作为公钥发布;严守机密的私钥是(2668, 157)。

对明文信息(英文)I love Zhejiang University 进行加密。

将这段文字转换为数字,不计大小写,每两个词之间为一个空格符号,空格符号对应数字 00,每个英文字母对应表征其在字母表中位置的两位数字,如:A 对应 01,B 对应 02,…,Z 对应 26,等等。再从前向后,将每四位数字划归一组,不足时补充空格。如此得到以下 13 组数字:

0900 1215 2205 0026 0805 1009 0114

0700 2114 0922 0518 1909 2025

每一组数字视为一个数,用公钥(2773, 17)对其加以变换。以第一个数为例,由于 $n=2773$,比任何可能出现的四位数字都大,故只需计算任何数字在模 2773 下的 17 次幂。所以

$$(900)^{17}=((((900)^2)^2)^2)^2\cdot 900\equiv 1510(\text{mod } 2773)$$

计算过程中,要随时注意取模,这样可以减少计算量。已求得 900 对应的密码为 1510。用同样的方法求得的全部密文电码为:

1510　0417　1524　1445　0542　2692　1684

0761　1644　2488　1787　1877　1672

解密过程与此类似,只不过使用密钥(2668, 157),直接计算虽很繁琐,但用计算机处理这一问题并不困难。本例中将四位数字划分为一组,是为了使每组的数字不超过 $n=2773$(根据事先规定,可能出现的最大数字为 2525)。当使用更大的 n 时,每次完全可以处理一个位数更多的数码组,只要相应的整数小于 n。

从代数结构角度来看 RSA 密码体制:RSA 使用了两个代数结构,一个是交换环 $R=\langle Z_n,+,\times\rangle$或交换半群$\langle Z_n,\times,1\rangle$,这个环或半群是公开的,用于加密和解密;一个是乘法群 $G=\langle Z^*_{\varphi(n)},\times\rangle$,用于生成密钥,这是一个保密的群。

9.6 密码学其他案例

9.6.1 符号密码求解

数学的应用是非常广泛的,有个译密码的问题也需要数学知识来解答。

设有密码 3・BIDFOR＝4・FORBID,其中,每个字表示一个十进制数字,试将这个密码译成数字的形式。

密码中含有 6 个未知数字,若将其逐个求出,实在困难,但稍微加以观察可知,只要将 BID,FOR 分别作为一个整体,就可将 6 个未知数字的问题转化为求两个数字的问题,从而可方便地解决。

设 $x=\text{BID}$,$y=\text{FOR}$,则

$$3(1000\,x+y)=4(1000\,y+x),$$

即得 $428x=571y$,

由于 x,y 都是三位数,且 428 与 571 互质,

所以 $x=\text{BID}=571$,$y=\text{FOR}=428$,

所以整个密码可译成 $3\times 571428=4\times 428571$。

9.6.2 魔阵密码的编译

我们将设计一种所谓魔矩阵,用它编发密码,且给出友方抄收后的翻译公式。

令 $N_n=\{1, 2, 3, \cdots, 2^n-1\}$, $n\in N$,向量 $T_n=(1, 2^1, 2^2, \cdots, 2^{n-1})$。

引理 1 $\forall n\in N$, $\forall \alpha\in N_n$,则存在唯一的 n 个分量的 0—1 向量 h_n,使得 $\alpha=T_n\cdot h_n$。

引理 2 若 $t_1, t_2, \cdots, t_m$ 是 T_n 的 m 个相异分量,$1\leqslant m\leqslant n$,则 $\sum\limits_{i=1}^{m}t_i\in N_n$.

定理 1 设 $S_j^{(n)}$ 是 0—1 魔阵 $W_{n\times 2^n}$ 的第 j 列向量,则成立公式 $j=T_n\cdot(S_j^{(n)})^{\mathrm{T}}$。

定理 2 $W_{n\times 2^n}$ 的列向量集合在 F_2 上构成一个 n 维线性空间 H_n。

有了以上的概念与结论,我们构造 $W_{5\times 2^5}$,且其第 1 列至第 26 列分别代表字母 $A, B, C, D, \cdots, X, Y, Z$;第 27 至第 30 列分别代表阴平、阳平、上声、去声的符号ˉ ˊ ˇ ˋ 。

魔阵 $W_{5\times 2^5}=$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0
0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Q	R	S	T	U	V	W	X	Y	Z	ˉ	ˊ	ˇ	ˋ	
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

下面给出利用 $W_{n\times 2^n}$ 的密码编译方法。

设 $a_1, a_2, \cdots, a_k$ 是 k 种事物,取 $n_0=\min\{n\mid k\leqslant 2^n-1\}$,构造 $W_{n_0\times 2^{n_0}}$,用 $W_{n_0\times 2^{n_0}}$ 的第 j 列 $S_j^{(n_0)}$ 代表事物 a_j,$1\leqslant j\leqslant k$,则用 $W_{n_0\times 2^{n_0}}$ 可以把以 $a_i(1\leqslant i\leqslant k)$为项的任意有穷序列用 0—1 码抄出,再代入公式

$$j=T_{n_0}\cdot(S_j^{(n_0)})^{\mathrm{T}},$$

而求得这一序列。

例如,$a_1, a_2, \cdots, a_{26}$ 分别是 $A, B, C, D, \cdots, X, Y, Z$,$a_{27}=$ˉ,$a_{28}=$ˊ,$a_{29}=$ˇ,$a_{30}=$ˋ,则

$$n_0=\min\{n\mid 30\leqslant 2^n-1\}=5,$$

于是我们采用 $W_{5\times 2^5}$ 来编译密码。

若抄收的密码是：

0 0 1 1 0 1 0 0 1 0 1 0 1 0 1 1 0 1
1 1 1 0 0 0 0 0 1 1 1 0 0 1 1 1 1 0
0 0 1 0 1 0 1 0 1 1 0 0 0 0 1 1 0 1
1 1 0 1 1 0 1 0 0 1 0 0 1 1 1 0 1 1
1 0 0 0 0 1 1 1 1 0 0 1 1 1 1 1 1 0
1 0 1 0 1 0 1 1 1 1 1 0 0 1 1 1 0 0
1 0 0 1 1 1 1 1 1 0 0 0 1 0 1 0 1 0
1 1 1 0 1 1 0 1 1 .

翻译公式为 $T_5 \cdot x_i=(1, 2, 4, 8, 16)\cdot x_i=j=a_j$，
其中 x_i 是抄收码中“第 i 段”向量，每段 5 个数码，所以得

$T_5 \cdot x_1=(1, 2, 4, 8, 16)\cdot(0, 0, 1, 1, 0)=4+8=12=L$,

$T_5 \cdot x_2=(1, 2, 4, 8, 16)\cdot(1, 0, 0, 1, 0)=1+8=9=I$,

$T_5 \cdot x_3=(1, 2, 4, 8, 16)\cdot(1, 0, 1, 0, 1)=1+4+16=21=U$,

$T_5 \cdot x_4=(1, 2, 4, 8, 16)\cdot(1, 0, 1, 1, 1)=1+4+8+16=29=$ ˇ,

$T_5 \cdot x_5=(1, 2, 4, 8, 16)\cdot(1, 0, 0, 0, 0)=1=A$,

$T_5 \cdot x_6=(1, 2, 4, 8, 16)\cdot(0, 1, 1, 1, 0)=2+4+8=14=N$,

$T_5 \cdot x_7=(1, 2, 4, 8, 16)\cdot(0, 1, 1, 1, 1)=2+4+8+16=30=$ ˋ,

$T_5 \cdot x_8=(1, 2, 4, 8, 16)\cdot(0, 0, 0, 1, 0)=8=H$,

$T_5 \cdot x_9=(1, 2, 4, 8, 16)\cdot(1, 0, 1, 0, 1)=1+4+16=21=U$,

$T_5 \cdot x_{10}=(1, 2, 4, 8, 16)\cdot(1, 0, 0, 0, 0)=1=A$,

$T_5 \cdot x_{11}=(1, 2, 4, 8, 16)\cdot(1, 1, 0, 1, 1)=1+2+8+16=27=$ ˉ,

$T_5 \cdot x_{12}=(1, 2, 4, 8, 16)\cdot(1, 0, 1, 1, 0)=1+4+8=13=M$,

$T_5 \cdot x_{13}=(1, 2, 4, 8, 16)\cdot(1, 0, 0, 1, 0)=1+8=9=I$,

$T_5 \cdot x_{14}=(1, 2, 4, 8, 16)\cdot(0, 1, 1, 1, 0)=2+4+8=14=N$,

$T_5 \cdot x_{15}=(1, 2, 4, 8, 16)\cdot(1, 1, 1, 0, 0)=1+2+4=7=G$,

$T_5 \cdot x_{16}=(1, 2, 4, 8, 16)\cdot(0, 0, 1, 1, 1)=4+8+16=28=$ ˊ,

$T_5 \cdot x_{17}=(1, 2, 4, 8, 16)\cdot(1, 0, 0, 1, 1)=1+8+16=25=Y$,

$T_5 \cdot x_{18}=(1, 2, 4, 8, 16)\cdot(1, 1, 1, 1, 0)=1+2+4+8=15=O$,

$T_5 \cdot x_{19}=(1, 2, 4, 8, 16)\cdot(1, 0, 1, 0, 1)=1+4+16=21=U$,

$T_5 \cdot x_{20}=(1, 2, 4, 8, 16)\cdot(0, 1, 1, 1, 1)=2+4+8+16=30=$ ˋ,

$T_5 \cdot x_{21}=(1, 2, 4, 8, 16)\cdot(1, 0, 0, 1, 1)=1+8+16=25=Y$,

$T_5 \cdot x_{22}=(1, 2, 4, 8, 16)\cdot(1, 0, 0, 1, 0)=1+8=9=I$,

$T_5 \cdot x_{23}=(1, 2, 4, 8, 16)\cdot(0, 1, 1, 1, 1)=2+4+8+16=30=$ ˋ,

$T_5 \cdot x_{24} = (1, 2, 4, 8, 16) \cdot (1, 1, 0, 0, 0) = 1+2=3=C$,
$T_5 \cdot x_{25} = (1, 2, 4, 8, 16) \cdot (1, 0, 1, 0, 1) = 1+4+16=21=U$,
$T_5 \cdot x_{26} = (1, 2, 4, 8, 16) \cdot (0, 1, 1, 1, 0) = 2+4+8=14=N$,
$T_5 \cdot x_{27} = (1, 2, 4, 8, 16) \cdot (1, 1, 0, 1, 1) = 1+2+8+16=27=$ ˉ 。

以上 0—1 密码译成拉丁拼音句子如下：

Liǔànhuāmíngyòuyìcūn

汉语译文:柳暗花明又一村。

9.6.3 选举问题

如何选举？怎样选举才公平？前一个是领导决策的问题,即指定选举的方法;而后一个需要精确的计算,是数学家要解决的问题。怎样在公平的意义下给出诸多政策的一个排序,是选举理论研究的问题。遗憾的是,要寻找一个完美公平的选举机制几乎是不可能的。试看下面的例子。

内阁 55 位成员要在 5 个党派的提名代表(记为 A,B,C,D,E)中确定一位担任总统。现在要求每个内阁成员按他(她)们的选择对 5 位候选人进行排序,结果如下。

内阁成员	18	12	10	9	4	2
第一选择	A	B	C	D	E	E
第二选择	D	E	B	C	B	C
第三选择	E	D	E	E	D	D
第四选择	C	C	D	B	C	B
第五选择	B	A	A	A	A	A

假定每位内阁成员都严格实施真诚选举,即始终以个人偏爱的选择进行表决,究竟谁能当上总统呢？通常有下列方法。

方法一,多数原则。候选人 A 以最多的 18 票当选。尽管得票数不足内阁成员的 1/3,但毕竟第一拥护者最多。

方法二,两轮选举。第一轮"海选"得票前两名进行第二轮选举,结果 B 以 37 票当选。

方法三,逐轮选举。类似奥林匹克运动会主办城市的选拔,每轮淘汰一名得票最少的。容易算出被淘汰的依次是 E、D、B、A,最后 C 当选。

方法四,博尔达计分法。以 18 世纪法国数学家博尔达(J. C. Borda,

1733—1799)名字命名的方法。如以5分、4分……记第一选择、第二选择……则可计算出各提名候选人的得分。A:127分;B:156分;C:162分;D:191分;E:189分。因此D以最高分当选。

方法五,鹰派对决。每两个候选人之间进行表决,则E∶A=37∶18;E∶B=33∶22;E∶C=36∶19;E∶D=28∶27。E当选。

分别看每一种选举方法,都有一定道理,但结果却是人人都可以当总统,没了道理。如果再加上策略选举或非真诚选举,情况更为复杂。1951年美国数学家阿罗(K. J. Arrow,1921—)发表一个"不可能性定理":绝对公平的选举系统是不存在的。他列出了5个公认的选举性质,证明了在任何一种情况下,都无法找到一个选举方法同时满足这些性质。

附:非真诚选举实例,2005年7月6日确定2012年奥林匹克夏季运动会主办城市的投票结果:

城市	第一轮	第二轮	第三轮	第四轮
伦敦	22	27	39	54
巴黎	21	25	33	50
马德里	20	32	31	
纽约	19	16		
莫斯科	15			

注意,在第一轮纽约得19票,但第二轮只得了16票,说明至少有3个投票人在第一轮"不真诚"。据说是为了照顾美国人的面子,使纽约不要在第一轮就被淘汰。还要注意马德里在第二轮的32票,而第三轮得31票,显然也是不真诚选举的实例。不过2005年12月爆出新闻:一位希腊奥委会投票人宣称,他在第三轮按错了表决器,把本该给马德里的票给了巴黎。马德里闻讯后推测:如果这样,马德里和巴黎在第三轮票数相同,国际奥委会主席罗格的决定性一票很可能投给马德里,而马德里在与伦敦的竞争中也会占上风,继而最终2012年夏季奥运会的主办地应该是马德里。可见,实际的选举更为复杂。

思考题

1. 第二次世界大战期间,美军破译了日军密码,保证了在太平洋战争中获胜,同时靠埋伏击落日海军大将山本五十六座机。请查阅相关史料,概述日军密码体系。

2. 第二次世界大战中的英德密码大战，主要是围绕着著名的Enigma(恩尼格码)展开的。二战中德国使用的密码是Enigma密码——十一种滚轴逻辑编辑密码。由波兰密码专家破译。由于德军入侵，波兰将密码的破译方法转送英国，英国是二战中破译德国密码最好的国家。德国海军使用的黑鲨密码，其复杂程度远超恩尼格码。英美密码破译部门费尽心机，最多只能测出德军潜艇发报时的位置而间接地推断出它的航线，或者根据德军的偶尔疏漏找出德军密码编译上的一些蛛丝马迹，但是一直到战后的一段时间，黑鲨都是盟军的心头大患，一直没有得到完全的破解。请查阅相关史料，概述黑鲨密码体系。

3. 假设要发送"李四将于周三17点45分到达16铺码头"的信息，请按不同方式编写所要发送的明文对应的密文。(1)使用密钥的单表代替加密，密钥Beijing；(2)仿射加密，$(a,b)=(7,3)$；(3)多表代替加密Hill(希尔)密码，K取例9.5的矩阵。

第10章 股市实战——预测与决策

股市是个开放的聚宝盆，智慧者可以从中尽情拾取财富；但它同时又像是魔鬼的藏宝洞，平庸者总是被关在其中永远不能解脱。投资者要想在股市里获利并赚取尽可能大的利润，就需要对股价未来运行的趋势作出尽可能准确的预测，并在实战时果断作出决策。

10.1 股票和股票交易

10.1.1 股票概念

股票是股份有限公司在筹集资本时向出资人发行的股份凭证。股票代表着其持有者(即股东)对股份公司的所有权。这种所有权是一种综合权利，如参加股东大会、投票表决、参与公司的重大决策等，收取股息或分享红利等。同一类别的每一份股票所代表的公司所有权是相等的。每个股东所拥有的公司所有权份额的大小，取决于其持有的股票数量占公司总股本的比重。股票一般可以通过买卖方式有偿转让，股东能通过股票转让收回其投资，但不能要求公司返还其出资。股东与公司之间的关系不是债权债务关系。股东是公司的所有者，以其出资额为限对公司负有限责任，承担风险，分享收益。

作为人类文明的成果，股份制和股票也适用于我国社会主义市场经济。企业可以通过向社会公开发行股票筹集资金用于生产经营。国家可通过控制多数股权的方式，用同样的资金控制更多的资源。

(1)股票基本特征和性质

①不可偿还性。股票是一种无偿还期限的有价证券，投资者认购了股票后，就不能再要求退股，只能到二级市场卖给第三者。股票的转让只意味着公司股东的改变，并不减少公司资本。从期限上看，只要公司存在，它所发行的股票就存在，股票的期限等于公司存续的期限。

②参与性。股东有权出席股东大会，选举公司董事会，参与公司重大决策。股票持有者的投资意志和享有的经济利益，通常是通过行使股东参与权

来实现的。股东参与公司决策的权利大小，取决于其所持有的股份的多少。从实践中看，只要股东持有的股票数量达到左右决策结果所需的实际多数时，就能掌握公司的决策控制权。

③收益性。股东凭其持有的股票，有权从公司领取股息或红利，获取投资的收益。股息或红利的大小，主要取决于公司的盈利水平和公司的盈利分配政策。

股票持有者凭股票定期从股份公司取得的收入是股息。股票只是对一个股份公司拥有的实际资本的所有权证书，只是代表取得收益的权利，是对未来收益的支取凭证，它本身不是实际资本，而只是间接地反映了实际资本运动的状况，从而表现为一种虚拟资本。

(2)股票的起源

股票至今已有400年的历史。股票是社会化大生产的产物。随着人类社会进入社会化大生产的时期，企业经营规模扩大与资本需求不足的矛盾日益突出，于是产生了以股份公司形态出现的、股东共同出资经营的企业组织；股份公司的变化和发展产生了股票形态的融资活动；股票融资的发展产生了股票交易的需求；股票的交易需求促成了股票市场的形成和发展；而股票市场的发展最终又促进了股票融资活动和股份公司的完善和发展。所以，股份公司、股票融资和股票市场的相互联系和相互作用，推动着股份公司、股票融资和股票市场的共同发展。

股票最早出现于资本主义国家。在17世纪初，随着资本主义大工业的发展，企业生产经营规模不断扩大，由此产生的资本短缺、资本不足便成为制约资本主义企业经营和发展的重要因素之一。为了筹集更多的资本，出现了以股份公司形态、由股东共同出资经营的企业组织，进而又将筹集资本的范围扩展至社会，产生了以股票这种表示投资者投资入股并按出资额的大小享受一定的权益和承担一定责任的有价凭证，并向社会公开发行，以吸收和集中那些分散在社会上的资金。1602年在荷兰成立的东印度公司是世界上最早的股份有限公司。股份有限公司这种企业组织形态出现以后，很快为资本主义国家广泛利用，成为资本主义国家企业组织的重要形式之一。伴随着股份公司的诞生和发展，以股票形式集资入股的方式也得到发展，并且产生了买卖交易转让股票的需求。这样，就带动了股票市场的出现和形成，并促使股票市场完善和发展。据文献记载，早在1611年就曾有一些商人在荷兰的阿姆斯特丹进行荷兰东印度公司的股票买卖交易，形成了世界上第一个股票市场，即股票交易所。目前，股份有限公司已经成为资本主义国家最基本的企业组织形式；股票已经成为资本主义国家企业筹资的重要渠道和方式，亦是

投资者投资的基本选择方式;而股票的发行和市场交易亦已成为资本主义国家证券市场的重要基本经营内容,成为证券市场不可缺少的重要组成部分。

10.1.2 股票交易

(1)股票交易场所

股票交易是指股票投资者之间按照市场价格对已发行上市的股票所进行的买卖。股票公开转让的场所首先是证券交易所。股票市场也称为二级市场或次级市场,是股票发行和流通的场所,也可以说是指对已发行的股票进行买卖和转让的场所。股票的交易都是通过股票市场来实现的。一般地,股票市场可以分为一级和二级市场,一级市场也称为股票发行市场,二级市场也称为股票交易市场。股票是一种有价证券。有价证券除股票外,还包括国家债券、公司债券、不动产抵押债券等。国家债券出现较早,是最先投入交易的有价债券。随着商品经济的发展,后来才逐渐出现股票等有价债券。因此,股票交易只是有价债券交易的一个组成部分,股票市场也只是多种有价债券市场中的一种。目前,很少有单一的股票市场,股票市场不过是证券市场中专营股票的地方。

证券交易所是依据国家有关法律,经政府证券主管机关批准设立的集中进行证券交易的有形场所。在中国有两个:上海证券交易所和深圳证券交易所。证券交易所分为公司制和会员制两种。这两种证券交易所均可以是政府或公共团体出资经营的(称为公营制证券交易所),也可以是私人出资经营的(称为民营制证券交易所),还可以是政府与私人共同出资经营的(称为公私合营的证券交易所)。

股票发行是指符合条件的发行人按照法定的程序,向投资人出售股份、募集资金的过程。

股份公司发行的股票,在经有关部门批准后,就可以在股票市场(证券交易所)公开挂牌进行上市交易活动。股票要上市交易必须具备一定的条件,并按一定的原则和程序进行操作与运转。

(2)股票交易名词

①股票价格:股票当做商品出卖时成交的价格。股票本身没有价值,股票价格通常不等于股票票面的金额。股票的票面额代表投资人股的货币资本数额,它是固定不变的;而股票价格则是变动的,它经常是大于或小于股票的票面金额。股票的买卖实际上是买卖获得股息的权利,因此股票价格不是它所代表的实际资本价值的货币表现,而是一种资本化的收入。股票价格一般是由股息和利息率两个因素决定的。例如,有一张票面额为100元的股票,

每年能够取得10元股息，即10%的股息，而当时的利息率只有5%，那么，这张股票的价格就是10元÷5%=200元。计算公式是：

股票价格=股息/利息率

可见，股票价格与股息成正比例变化，而和利息率成反比例变化。如果某个股份公司的营业情况好，股息增多或是预期的股息将要增加，这个股份公司的股票价格就会上涨，反之，则会下跌。对于股息的不同预期是股票价格波动的最主要原动力。

②开盘价：当日开盘后该股票的第一笔交易成交的价格。如果开市后30分钟内无成交价，则以前日的收盘价作为开盘价。

③收盘价：每天成交中最后一笔股票的价格，也就是收盘价格。

④最高价：当日成交的价格中的最高价位。

⑤最低价：当日成交的价格中的最低价位。

⑥成交量：在某一时段内具体的交易数。它可以在分时图中绘制，包括在日线图、周线图、月线图，甚至是5分钟、30分钟、60分钟图中绘制。市场成交量的变化反映了资金进出市场的情况，成交量是判断市场走势的重要指标。一般情况下，成交量大且价格上涨的股票，趋势向好。成交量持续低迷时，一般出现在熊市或股票整理阶段，市场交易不活跃。成交量是判断股票走势的重要依据，对分析主力行为提供了重要的依据。投资者对成交量异常波动的股票应当密切关注。

⑦内盘：以买入价成交的交易，买入成交数量统计加入内盘。

⑧外盘：以卖出价成交的交易，卖出成交数量统计加入外盘。

内盘、外盘这两个数据大体可以用来判断买卖力量的强弱。若外盘数量大于内盘，则表现为买方力量较强，若内盘数量大于外盘则说明卖方力量较强。通过外盘、内盘数量的大小和比例，投资者通常可能发现主动性的买盘多还是主动性的抛盘多，并在很多时候可以发现庄家动向，是一个较有效的短线指标。但投资者在使用外盘和内盘时，要注意结合股价在低位、中位和高位的成交情况以及该股的总成交量情况。因为外盘、内盘的数量并不是在所有时间都有效，在许多时候外盘大，股价并不一定上涨；内盘大，股价也并不一定下跌。

⑨股票指数：由证券交易所或金融服务机构编制的反映不同时点上股价变动情况的相对指标。由于股票价格起伏无常，投资者必然面临市场价格风险。对于具体某一种股票的价格变化，投资者容易了解，而对于多种股票的价格变化，要逐一了解，既不容易，也不胜其烦。为了适应这种情况和需要，一些金融服务机构就利用自己的业务知识和熟悉市场的优势，编制出股票价

格指数，公开发布，作为市场价格变动的指标。投资者据此就可以检验自己投资的效果，并用以预测股票市场的动向。同时，新闻界人士、公司老板乃至政界领导人等也以此为参考指标，来观察、预测社会政治、经济发展形势。通常是将报告期的股票价格与约定的基期价格相比，并将两者的比值乘以基期的指数值定义为该报告期的股票指数。股票指数的计算方法有三种：一是相对法，二是综合法，三是加权法。

⑩世界上几种著名的股票指数：美国的道琼斯、标准普尔、纳斯达克三大股票指数，日本的日经道琼斯股价指数，英国的伦敦金融时报股票价格指数，中国香港的恒生指数，中国内地的上证综指和深圳成指以及沪深 300 指数等。

10.2 波段操作

俗话说“收益与风险并存”，要想获取比别人更高的利润必然要冒更大的风险。证券市场就是冒险的乐园，也是普通老百姓实现发财梦想的沃土。每个大型证券市场至少都有几百万投资者，大家都想赚钱，多空必然经常明显地存在，股票价格波动似乎难以捉摸。但是国内外市场运行结果显示，其中又似乎有一定规律。这就有了各种技术分析和所谓的理论，其中最著名的当属波浪理论。

10.2.1 股市的波浪理论

美国人艾略特(R. N. Elliot，1871—1948)将股市的上升、下降的波动趋势形象地表示为大海的波浪，一浪推一浪，潮起潮落。如能把握住股市的波浪起伏，就能够顺应潮流。中国股市的实际情况通常有如下几点。

(1)上升态势浪一般由五个浪型组成。第一浪为启动浪；第二浪为初次调整浪；第三浪为发展浪；第四浪为再次调整浪；第五浪为冲高浪。

(2)下跌态势浪一般由三浪组成。A 浪为下跌出货浪；B 浪为反弹出货浪；C 浪为出货探底浪。

(3)从浪形构图观察：全五浪的上升态势浪和全三浪的下跌态势浪完整构成一个股市潮起潮落的态势图。

(4)波浪理论只是根据波浪起伏形象地描述股市的升降起落，只是为炒股提供了一个参考工具而已，并不是说股市就固定不变地、机械地按事先框好的浪型、浪顶、浪底、时间、斜率等按部就班地运作，实际上经常有变化。

(5)虽然波浪理论的基本框架是由上五浪和下三浪构成，但在具体每一个浪型中，还会出现子浪、孙浪，而分得过细又反而把人搞得满头雾水。因

此，只要大体能判断出大波浪的运行方向就可以了。

纵观中外股市，局部K线走势基本上都能归并为三种类型。

(1)横盘整理。一幅度范围内上下来回振荡，后一个高点可能超过前一个高点，或者后一个低点低于前一个低点，高于或低于的幅度较小，也可能不超过。该阶段是因为投资者们分歧较大，造成多空拉锯；整理到一方的力量明显增大为止。

(2)单边上涨或者单边下跌。K线图呈现明显的至少45度角上涨通道，或明显的至少45度角下降通道。该阶段是因为投资者们分歧较小，行情朝着与多数人一致的方向运行，行情通常伴随着越来越多人开始改变主意而戛然而止。

显然，如何正确地数浪是关键，本书介绍一种笔者发现的相当有效的有序波浪聚类方法。

10.2.2 聚类分析

聚类是将物理或抽象对象的集合分成若干由类似的对象组成的多个类的过程。由聚类生成的簇是一组数据对象的集合，这些对象与同一个簇中的对象彼此相似，与其他簇中的对象相异。现在聚类分析广泛应用于生物、经济、社会、人口等领域的大量量化分类问题的研究中，也是数据处理、数据挖掘中最主要的内容之一。聚类分析内容非常丰富，有系统聚类法、动态聚类(K均值聚类)法、有序样品聚类法和模糊聚类法等。

在实际应用中，有时样品之间有先后次序且不能变动，在同一类中的样品要求次序相邻的，这类问题称为有序样品的聚类分析。例如对动植物按生长的年龄段进行分类，年龄的顺序是不能改变的，否则就没有实际意义了；在地质勘探中，需要通过岩心了解地层结构，此时按深度顺序取样，样品的次序也不能打乱；在证券交易中，对行情的研判必须与交易时间相对应，今天赚钱的品种在昨天或者明天都可能严重亏损。

10.2.3 上证综指的聚类

表10-1是2010年7月16日到2011年7月15日整一年共243个交易日的上证综指收盘价(来自同花顺交易软件)，图10-1是相应的散点折线图。

表 10-1　243 个交易日的上证综指收盘价

日期	收盘价	日期	收盘价	日期	收盘价	日期	收盘价
2010-07-16	2424.27	2010-10-21	2983.53	2011-01-17	2706.66	2011-04-21	3026.67
2010-07-19	2475.42	2010-10-22	2975.04	2011-01-18	2708.98	2011-04-22	3010.52
2010-07-20	2528.73	2010-10-25	3051.42	2011-01-19	2758.1	2011-04-25	2964.95
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2010-10-19	3001.85	2011-01-13	2827.71	2011-04-19	2999.04	2011-07-15	2820.17
2010-10-20	3003.95	2011-01-14	2791.34	2011-04-20	3007.04		

近几年来，中国经济虽然保持每年 9%左右的增长，但是自金融危机以来整个国际大环境并不是特别地明朗，再加上持续扩容和宏观调控的影响，中国股市处于上下两难的尴尬境地。很明显，温和的振幅不大的横盘整理代替了原来的快速力度大的反弹或回调，横盘箱体上下幅度一般不超过 5%；如果幅度超过 5%，则相应地变成上涨或下降。

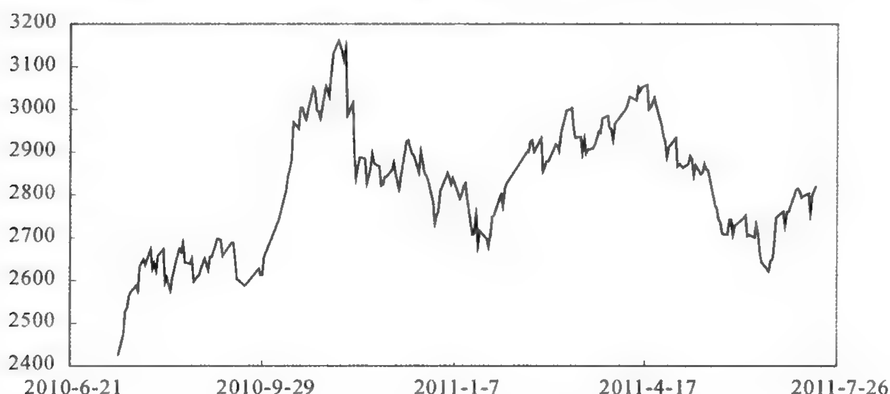


图 10-1　243 个交易日的上证综指收盘价折线图

针对上证综指，构造有序波段聚类算法。

(1)在上升波段的某一高点 A(所谓高点指前一天与后一天都比该天低)后的 3 周 15 个交易日内，若出现超出 2%以上的新高点 B，则上涨行情延续，B 取代 A；若不能出现超出 2%以上的新高点而且出现 5%以上的下跌低点，则上涨行情结束，其中的最高点作为上涨行情的结束日，最高点的次日作为行情转入下跌走势的起点；若不能出现超出 2%以上的新高点也不出现 5%以上的下跌低点，则上涨行情结束，高点 A 就是其最后一天，A 后的下一天作为进入横盘振荡的首日。下降波段完全类似。

(2)设置横盘振荡(或称箱体振荡整理)幅度限5%,若某交易日收盘价高出横盘时最低收盘价的5%,则意味着箱体振荡结束,新的上涨行情开始;同样地,若某交易日收盘价低于横盘时最高收盘价的5%,则意味着箱体振荡结束,新的下跌行情开始。

(3)运算按时间从最早日子开始,每个样品都只扫描一次,与当前类G的最大值、最小值进行比较,重复(1)和(2)的过程,判断当前类是否结束。

注意,针对不同数据对象,其中的百分比准则要作相应调整。如2006—2007年大牛市中上证综指箱体振荡至少10%以上,甚至高达20%。

根据有序波段聚类算法的运算结果见表10-2,其中涨跌幅计算公式如下:上涨时的涨幅=极差/最低值,下跌时的跌幅=极差/最高值,横盘时的振幅=2极差/(最低值+最高值),就是最大涨跌幅和最大振幅。应该说,运算结果令人满意,运算速度明显优于Fisher最优分割法。

表10-2 243个交易日的上证综指有序聚类结果

时间	波段	最高	最低	极差	涨跌幅(%)	交易日数
2010-07-16 至 2010-08-02	上涨	2672.52	2424.27	248.25	10.24	12
2010-08-03 至 2010-09-30	横盘	2698.36	2575.47	122.89	4.66	40
2010-10-08 至 2010-11-08	上涨	3159.51	2738.74	420.77	15.36	22
2010-11-09 至 2010-11-17	下跌	3147.74	2838.86	308.88	10.88	7
2010-11-18 至 2010-12-27	横盘	2927.08	2781.40	145.68	5.10	28
2010-12-28 至 2011-01-25	下跌	2838.80	2677.43	161.37	5.68	20
2011-01-26 至 2011-03-09	上涨	3002.15	2708.81	293.34	10.83	26
2011-03-10 至 2011-04-27	横盘	3057.33	2925.41	131.92	4.41	33
2011-04-28 至 2011-06-20	下跌	2932.19	2621.25	310.94	10.60	36
2011-06-21 至 2011-07-15	上涨	2820.17	2646.48	173.69	6.56	19

10.3 技术指标分析

10.3.1 技术分析与基本分析

股票技术指标是相对于基本分析而言的。基本分析法着重于对一般经济情况以及各个公司的经营管理状况、行业动态等因素进行分析,以此来研究股票的价值,衡量股价的高低。而技术分析则是透过图表或技术指标的记录,研究市场过去及现在的行为反应,以推测未来价格的变动趋势。其依据

的技术指标的主要内容是由股价、成交量或涨跌指数等数据计算而得的，我们也由此可知，技术分析只关心证券市场本身的变化，而不考虑会对其产生某种影响的经济方面、政治方面等的各种外部的因素。

基本分析的目的是为了判断股票现行股价的价位是否合理并描绘出它长远的发展空间，而技术分析主要是预测短期内股价涨跌的趋势。通过基本分析我们可以了解应购买何种股票，而技术分析则让我们把握具体购买的时机。在时间上，技术分析法注重短期分析，在预测旧趋势结束和新趋势开始方面优于基本分析法，但在预测较长期趋势方面则不如后者。大多数成功的股票投资者都是把两种分析方法结合起来加以运用。他们用基本分析法估计较长期趋势，而用技术分析法判断短期走势和确定买卖的时机。

股价技术分析和基本分析都认为股价是由供求关系决定的。基本分析主要是根据对影响供需关系种种因素的分析来预测股价走势，而技术分析则是根据股价本身的变化来预测股价走势。技术分析的基本观点是：所有股票的实际供需量及其背后起引导作用的种种因素，包括股票市场上每个人对未来的希望、担心、恐惧等，都集中反映在股票的价格和交易量上。

技术分析理论基础：①市场行为包容消化一切影响价格的任何因素，基本面、政治因素、心理因素等都要最终通过买卖反映在价格中，也就是价格变化反映供求关系，供求关系决定价格变化；②价格以趋势方式演变对于已经形成的趋势来讲，通常是沿现存趋势继续演变；③历史会重演技术分析和市场行为学与人类心理学有一定关系，价格形态通过特定的图表表示了人们对某市场看好或看淡的心理。

空中楼阁理论：空中楼阁理论是美国著名经济学家凯恩斯于 1936 年提出的，该理论完全抛开股票的内在价值，强调心理构造出来的空中楼阁。投资者之所以要以一定的价格购买某种股票，是因为他相信有人将以更高的价格向他购买这种股票。至于股价的高低，这并不重要，重要的是存在更大的“笨蛋”愿以更高的价格向你购买。精明的投资者无需去计算股票的内在价值，他要做的只是抢在最大“笨蛋”之前成交，即股价达到最高点之前买进股票，而在股价达到最高点之后将其卖出。

10.3.2 技术分析分类

(1)机械趋势交易

交易者的买与卖完全死板(如机械化)，当某一多头信号出现时，即进场买进；空头信号出现时，则进场卖出。这种类型利用的图表工具有 K 线图技术、移动平均线、RSI 线、KD 线等。趋势分析的原理是建立在多空头双方的

决策比较上,比如多头趋势,当股价回落到趋势线上时即为买入信号,这时前期的多头波段累积的量能对股价有一定的支撑作用,这是空间法则的表现。

(2)线路趋势交易

利用这种工具是指以线路的技术性知识来决定市场的交易,必须时时注意长期的趋势线路(周线或月线)的走势,找出市场汇价的支持位和阻力位。

10.3.3 技术分析优点与缺点

(1)技术分析优点

技术分析具备全面、直接、准确、可操作性强、适用范围广等显著特点。与基本分析相比,技术分析进行交易见效快,获得利益的周期短。此外,技术分析对市场的反应比较直接,分析的结果也更接近实际市场的局部现象。相比较基本分析而言,通过市场分析得到的进出场位置往往比较准确。

(2)技术分析缺点

技术分析的缺点是考虑对象的范围相对较窄,对长远的市场趋势难以进行有效的判断。基本分析主要适用于周期相对比较长的市场预测以及预测精确度要求不高的领域。技术分析相对于基本分析更适用于短期的行情预测,要进行周期较长的分析则必须参考基本分析,这是应用技术分析最应该注意的问题。因为技术分析是经验的总结而非科学体系,所以通过技术分析得到的结论并由此进行的交易操作需要以概率的形式为投资者带来收益。

10.3.4 技术分析指标简介

(1)BIAS 乖离率

乖离率表现个股当日收盘价与移动平均线之间的差距。正的乖离率愈大,表示短期获利愈大,则获利回吐的可能性愈高;负的乖离率愈大,则空头回补的可能性愈高。按个股收盘价与不同天数的平均价之间的差距,可绘制不同的 BIAS 线。

(2)DMI 趋向指标

DMI 中文名称趋向指标,是一种相当常用的指标。其基本用法是看+DI 和-DI 交叉。+DI 由下往上交叉-DI,为买进信号;+DI 由上往下交叉-DI,为卖出信号。ADX 在 20 以下,代表股价处于整理期,应离场观望;ADX 突破 20～30 向上爬升,将有一段相当幅度的涨跌;ADX 高于 50 以上,忽然转弯向下反折,无论此时是上涨或下跌,都表明行情即将反转。

(3)EXPMA 指数平滑移动平均线

为了解决移动平均线落后的问题,分析学家另外寻求 EXPMA 均线指标

用于取代移动平均线。EXPMA 可以随股价的快速移动,立即调整方向,有效地解决讯号落后的问题。当第一条 EPMA 由下往上穿越第二条 EXPMA 时,将对股价造成推升力道。当第一条 EXPMA 由下往上穿越第二条 EXPMA 的,将对股价造成推降力道。股价由下往上碰触 EXPMA 时,很容易遭遇大压力回档。股价由上往下碰触 EXPMA 时,很容易遭遇大支撑反弹。

(4)KDJ 随机指标

K 值在 20 左右,向上交叉 D 值为短期买进信号。K 值在 80 左右,向下交叉 D 值为短期卖出信号。K 值形成一底比一底高的现象,并且在 50 以下的低水平,由下往上连续两次交叉 D 值时,股价涨幅会较大。K 值形成一顶比一顶低的现象,并且在 50 以上的高水平,由上往下连续两次交叉 D 值时,股价跌幅会较大。

(5)MACD 指数平滑异同平均线

该指标主要是利用长短期两条平滑平均线,计算两者之间的差离值。该指标可以去除掉移动平均线经常出现的假讯号,又保留了移动平均线的优点。但由于该指标对价格变动的灵敏度不高,属于中长线指标,所以在盘整行情中不适用。图中柱线由绿翻红是买入信号,由红翻绿是卖出信号。MACD 曲线由高档两次向下交叉时,则股价下跌幅度会较深。MACD 曲线由低档两次向上交叉时,则股价上涨幅度会较大。股价高点比前一次高点高,而 MACD 指标的高点却比前一次高点低时,为牛背离,暗示股价会反转下跌。

(6)ROC 变动速率

当 ROC 向下跌破零,卖出信号;ROC 向上突破零,买入信号。股价创新高,ROC 未配合上升,显示上涨动力减弱。股价创新低,ROC 未配合下降,显示下跌动力减弱。股价与 ROC 从低位同时上升,短期反弹有望。股价与 ROC 从高位同时下降,警惕回落。

(7)RSI 相对强弱指标

短期 RSI 在 20 以下水平,由下往上交叉长期 RSI,为买进信号。短期 RSI 在 80 以上水平,由上往下交叉长期 RSI,为卖出信号。股价一波比一波低,相反地,RSI 却一波比一波高时,股价很容易反转上涨。股价一波比一波高,相反地,RSI 却一波比一波低时,股价很容易反转下跌。RSI 在 50 以下为弱势区,50 以上为强势区。由下向上突破 50 线为由弱转强,由上向下突破 50 线为由强转弱。一般认为 RSI 在 50 以上准确性较高。

(8)TRIX 三重指标平滑平均线

该指标是一种三重指数平滑平均线,长线操作时采用本指标的讯号,可

以过滤掉一些短期波动的干扰,避免交易次数过于频繁,造成部分无利润的买卖及手续费的损失,但该指标在盘整行情中不适用。TRIX 向上交叉其 MA 线为买入讯号,TRIX 向下交叉其 MA 线为卖出讯号。有种看法认为该指标在判断卖出时可能会失真。

(9)W&R 威廉指标(William's %R)

低于 20,超买,即将见顶,应及时卖出。高于 80,超卖,即将见底,应伺机买进。这个指标与 RSI、MTM 指标配合使用,效果更好。

10.3.5 技术分析步骤

(1)看图

图表的类型很多,但是都大同小异。最基本和常用的是阴阳蜡烛图。一根柱子表示一个时间段。阴柱表明下跌,即收市价低于开市价;阳柱表明上涨,即收市价高于开市价。下两条边是该时间段的开盘价和收盘价,实体上部和下部的垂直线影线分别代表该时间段的最高价和最低价。

通过对蜡烛图的形态分析,主要可以分析行情什么时候将反转,在一波上升行情中判断出最早何时将调头向下,那么就可以第一时间做空赚取更大收益;同样在一波下跌行情中判断最早何时将企稳反弹,那么就可以第一时间做多占尽先机。蜡烛图还有一部分形态是研究行情在什么情况下会延续,如果掌握这些形态可以更好地顺势交易,持有有利单时也会更加充满信心地持有,甚至途中加仓来使收益更大。

(2)看趋势

技术分析的第二个关键,就是要学会画趋势线、阻力位、支撑位。支撑位是指存在较大支撑的价位,股价下跌到该价附近时容易企稳反弹;阻力位是指存在较大压力的价位,股价上升到该价位附近时容易遇阻回落。计算支撑和阻力位的方法主要用画趋势线的方法,也可以结合技术指标,如黄金分割位、均线系统、保利加通道等。

在上升趋势中,将两个上升的低点连成一条直线,就得到上升趋势线。在下降趋势中,将两个下降的高点连成一条直线,就得到下降趋势线。为了使画出的趋势线在今后分析市场走势时具有较高的准确性,我们要对用各种方法画出的试验性的趋势线进行筛选,去掉无用的,保留确实有效的趋势线。

要得到一条真正起作用的趋势线,要经多方面的验证才能最终确认,不合条件的一般应予以删除。

首先,必须确实有趋势存在。也就是说,在上升趋势中,必须确认出两个依次上升的低点;在下降趋势中,必须确认两个依次下降的高点,才能确认趋

势的存在，连接两个点的直线才有可能成为趋势线。其次，画出直线后，还应得到第三个点的验证才能确认这条趋势线是有效的。一般说来，画出的直线被触及的次数越多，其作为趋势线的有效性越被得到确认，用它进行预测越准确有效。此外，还要不断地修正原来的趋势线，例如当跌破上升趋势线后又迅速回升至这条趋势线上面，分析者就应该从第一个低点和最新形成的低点重画一条新线，或是从第二个低点和新低点修订出更有效的趋势线。

(3)结合技术指标，确认对市场趋势的判断

交易者在分析图表的时候，经常会运用一些技术分析指标，以更清楚地确认市场趋势，更好地掌握入市、出市点。关键是如何综合各个技术指标特征，因为技术指标经常会出现高位钝化时暂时失效，这时候需要具备操作的经验和果断性。

10.4 旁门左道

无论是处于绝对劣势的小散户，还是实力雄厚的基金操盘手，对后市行情趋势判断和热点把握都是取胜的关键。于是，有人将希望寄托在了各种“非常”的预测方法之中，比如把《周易》中的八卦运用于股市之中；穿红内裤避邪；在“章鱼哥”之后，“鳄鱼哥”又成了股民预测股市的新偶像……

10.4.1 隔夜美股测沪深开盘

当今是全球化时代，一个地区发生的事件很有可能会影响到相邻地区，甚至影响全球。金融市场尤为明显。美国作为世界霸主，一举一动都可能影响到世界的每一个角落。我们截取 2010 年 1 月 4 日至 12 月 8 日共 242 个交易日道琼斯指数与上海证券综合指数数据，其中美国有 5 日、中国有 16 日休市而另一方不休市。

先计算上海综合指数开盘涨幅＝100(今沪综开盘指数/昨沪综收盘指数－1)，表 10-3 列出部分数据。在这 10 对数据中，沪综指开盘涨跌与道指(道琼斯指数)昨收盘涨跌一致的有 7 天，不一致的有 3 天。统计同时开市的 221 的交易日，得沪综指开盘与道指昨收同步的共有 163 日，占总天数的 163/221＝73.76%，接近 3/4。再计算两者之间的简单相关系数得 0.679，两者线性相关性显著。因此，我们可以说，隔夜美股涨跌明显影响沪市开盘。进一步计算沪综指收盘涨跌与沪综指开盘涨跌及隔夜道指涨跌的相关系数分别是 0.334 和 0.231，远没有沪综指开盘与昨道指收盘关系密切。换句话说，沪综指收盘价受昨夜美股涨跌和自身开盘涨跌影响不大，即无法根据美股涨跌或

沪市开盘涨跌预测当天沪市涨跌。

表 10-3　2010 年 1 月 4 日—18 日道指收盘涨幅与次日沪综指开盘涨幅(%)

日期	4	5	6	7	8	11	12	13	14	15	18
道指收盘涨幅	1.50	−0.11	0.02	0.31	0.11	0.43	−0.34	0.50	0.28	−0.94	
沪综指开盘涨幅		0.33	−0.14	−0.01	−0.49	3.30	−0.22	−2.11	0.34	0.05	−0.33

相应地，用完全一样的方法研究沪市涨跌对当日美国股市开市的影响，因为中国股市收市早于美国股市开市。结果是 221 个交易日中，道指开盘涨跌与当日沪综收盘涨跌一致的共 124 日，占总天数的 124/221＝56.1%，明显小于 73.6%，而接近 50%；再计算两者之间的简单相关系数得 0.101，两者线性相关性不显著。即中国股市的涨跌对当天美国股市的影响几乎忽略不计。

10.4.2　主力持仓测后市

2010 年 8 月上旬，有人根据期指主力会员净持仓的变化，来研判股市后市走势，并曾连续 12 个交易日预测正确，因而这一指标也一度被市场称作股市的“章鱼哥”。所谓的“预测正确”是指期指主力合约前 20 名会员净持仓的变化方向与下一交易日的沪深 300 指数的涨跌方向一致。

不过，在连胜 12 局之后，“净持仓位”预测法后来则连连失手。对此，另有人指出，对比期指交易以来各个交易日的情况，主力前 20 位的持仓位的买卖单情况与第二天股市涨跌情况相符的几率只有 56%，因此“期指章鱼哥”的说法具有一定的偶然性。另外，由于期指目前的交易量与股市相比非常小，因此期指的各项指标与股市的联动都很难取得一致的步调，因此这些指标的走势也只能作为研判后市的参考，而不可画上等号。到底关联度如何，完全可采取类似上小节道指与沪指的计算方法进行计算分析，作为思考题留给读者自己查找数据并完成。

10.4.3　用周易八卦算涨跌

据报道，股民甲最近迷上了周易八卦，不过不是因为爱上了古代哲学，而是周易可以算出个股和大盘的走势。她买股之前，都要用八卦的占卜排序法先预测一下大盘的涨跌，然后再用同样的方法算一算要买的个股对应的“卦文”是否吉利或者是否能带来财运。

“任何事物都有阴阳对立面，有孕育、发展、鼎盛、衰败、病墓和死绝六个阶段，这是宇宙万物的普遍规律，周易的核心思想就是阐述了这个普遍规律，因此利用周易的理论来指导股票交易有其理论上的依据和联系。”甲表示，在

股票的曲线图上可以找到周易八卦中的规律。网络上甚至还出现了专门的算卦软件，只要输入个股的相关信息，立马就可以得到卦文以及相应的解答。

这里涉及两个问题，一是易经预测的合理性，二是如何预测。

首先，《易经》是科学吗？《易经》作为“群经之首，大道之源”，影响中国至少有两千多年历史，在中华传统文化中的地位非其他任何一本书所能比肩。“易”就是变易、变化莫测；《易经》就是变易之道，它试图对客观世界、自然现象、人文社会中诸现象用变易的思想进行解读。《易经》共有 64 卦，每一卦讲一件事。

以第一卦乾卦为例，从初爻到最后的上爻的爻词分别是：“①潜龙勿用；②见龙在田，利见大人；③君子终日乾乾，夕惕若，厉无咎；④或跃在渊，无咎；⑤飞龙在天，利见大人；⑥亢龙有悔。”将龙想象为人，以一个人的成长过程来解读之：①开始时你一无所有、什么也不会，除了学习，别指望别人会真正用你；②等到你掌握了点技术、有了点资本，似乎可以独立，但别急、多跟班，有机会多见见大人物，注意你还不是大人物；③从小人物变成大人物可不是件容易的事，你必须经常反省，今天哪儿做得好、哪儿做得不好，逐渐积累经验教训；④鲤鱼跳龙门，你只有跳才可能进龙门，你要主动出击多多表现，才有成功的机会，才有被大人物赏识的可能；⑤现在你是成功人士，高高在上，你要交往的自然应该是“人物”；⑥高处不胜寒，当你功德圆满之际，也是衰败的开始，请记住这个世界不是你一人的，要把财富功名留点给他人。夸张点说，易经中的每一卦都是放之四海而皆准的变化之道。如果以乾卦 6 爻预测股票，则分别解读为：①继续调整，观望；②刚脱离底部，仍有反复；③振荡整理，观望；④大幅振荡，快进快出；⑤主升浪，大胆跟进；⑥顶部，清仓。

事物的变易、变化当然是以时间为坐标，对每个人来说，时间都分为过去、现在和将来。每个人都是一样的，赤条条来到世上，最后衰老病死，不同的是甲生于帝王将相之家，乙生于穷人家。古往今来，芸芸众生，最后演绎出太多惊人相似的一幕幕。所以说利用《易经》能预测人事根本没问题。如果换个角度，周文王、孔夫子与我们相差约三千年，三千年对人类生物体进化几乎忽略不计，我们现代人只是比祖先们多些知识和工具，并不比他们更聪明。《易经》存在了三千年，必然存在相当的合理性。

其次，如何预测，才是关键。环境在变，社会在变，祖先们创造出的种种预测方法是否适合于今天（暂不去追究这些方法在过去是否合理）？我们无法回答！至少本书的主要作者回答不了，虽然他曾经潜心研究《易经》数年。《易经》预测包括起卦和解说，这两步都还问题多多。起卦方法有多种，最经典的是三枚铜钱抛 6 次起卦，后进化出数字起卦，这绝对是个伟大发明，可惜

到底用什么起数，如时间、像、物选择哪一个，如何转换成数？得到一个卦，如何解说？远不是某些书上写得那么轻松，这是最难的一关。因为篇幅，无法进一步展开探讨。

10.4.4 邪门

如果说易经预测有相当的理论作依据，接下来提及的实在难以找到什么理论，姑且称之为“邪门”。邪门招数：穿红内裤避邪、凭感觉、依梦境炒股、动物预测之类。

红内裤是本命年的象征，不过并非本命年的股民丙却买了数条红内裤。“最近亏得有点惨，穿条红内裤避避邪。”丙说，今年以来股市操作难度加大，想穿点红色的衣物来“转运”。

很多股民对与“牛”和“红”沾边的事物都很迷信。在语言上要避讳跌、亏、熊字眼及其同音字；在穿着上“穿红避绿”则是股民的共识。除此之外，狂吃牛肉，狂买“牛”饰也一度成为了股民们借以转运的手段。

此外，一些股民还经常把梦境和股市的涨跌联系起来。“前几天我做梦老是梦见发大水，结果这几天我的股票就一直下跌。”股民丁说，现在只要梦到发大水，他就绝对不买股票。

2008 年，俄罗斯一只名叫鲁莎的黑猩猩“亲自理财”，从 30 家上市公司中挑选出 8 家公司，鲁莎选择的股票翻了近 3 倍，从而在股市刮起了一波“猩猩预测风”。2010 年，世界杯袭来，章鱼保罗以 8 猜 8 中的完美预测能力赢得了全世界的掌声。“我以后再也不相信那些所谓的股神了，明天就到水产市场买一窝章鱼仔，每天让它们帮我选股票。”世界杯之后，股民戊在网上发帖，没想到引来了上千人的点击。“信章鱼得涨停！”一位跟帖的网友表示，他办公室炒股的同事目前都养起了章鱼。还有股民郁闷地表示，水产市场的大章鱼都卖断货了。“章鱼旋风”还没过，澳大利亚的选举又捧红了同样有预测“神力”的“鳄鱼哥”。不过这却让股民们犯起了难：喂鳄鱼的难度太大了，小散户们在股市赚的钱还不够鳄鱼的一顿饭钱。

实际上股市中真正的敌人是自己，缺乏操作纪律，随意性操作，战术战略弃之不顾，输给自己内心的贪婪和恐惧。

①贪婪。涨了希望再涨，技术已经见顶回调却指望反抽后再走，想要卖在最高价；大幅下跌后希望继续大跌，想要买到最低价；经常满仓操作，梦想一夜暴富。

②恐惧。下跌后担心还会下跌，割肉出局；机会出现后不敢介入，怕回调被套；大涨后担心还会涨，高位追高买入。

克服自身的“贪婪”和“恐惧”，坚持一定的技术分析原则，理性操作，就能成为证券市场中的赢家。

10.5 预测与决策

10.5.1 名词解释

从字面上解释，预测就是预先推测或测定，决策就是决定计策或办法。预测与决策都是针对未来之事，决策是最终目的，预测是辅助。

譬如张三想买房子，买还是不买，要作出选择。假如他原来没有房子，朋友已谈好，准备考虑找个时间结婚，必须有个自己的家，于是咬牙凑钱借钱也要买，这里的决策相对简单。如果张三已经有房子，只是考虑换一套大些或环境更好些的房子，而钱又不太宽裕，主要因为房价持续上涨，怕现有积蓄贬值。显然这时就不能草率决策，先要对不久的将来作些预测，房价和整个物价、自己的收入、其他投资理财等会变吗，朝哪个方向变？如果说预计未来房价比收入涨得快，那就出手买吧，当然经济实力要有一定保障。否则，暂时观察观察，过一段时间再作决定。

10.5.2 预测

(1)预测概念

“凡事预则立，不预则废”。“人无远虑，必有近忧”。很早以前，人们就试图预测未来。例如，对于出现的各种节气、日食或月食这类纯粹的天文地理现象，人们能凭借经验对其作出足够精确的预测。然而，处理当今错综复杂、不断发展变化的各种实际问题时，人们要作出决策就不那么容易了。因此，预测未来并不是靠想入非非或仅凭个人的冥思苦想，而是要靠科学，包括科学的头脑、科学的方法以及科学的手段等。预测就是根据研究对象发展变化的实际数据和历史资料，运用现代的科学理论和方法以及各种经验、判断和知识，对事物在未来一定时期内的可能变化情况进行预测，根据事物的过去和现在估计未来，根据已知预测未知，从而减少对未来事物认识的不确定性，以指导我们的决策行动，减少决策的盲目性。

由于预测对象、时间、范围、性质等的不同，可以有不同的预测方法分类。根据方法本身的性质特点，常用的预测方法分为三大类。

①定性预测方法。这类方法主要是依据人们对系统过去和现在的经验、判断和直觉，如市场调查、专家打分、主观评价等作出预测。主要有德尔斐法、主观概率法和领先指标法等。

②回归分析和时间序列分析预测方法。这类方法主要是根据系统对象随时间变化的历史资料(如统计数据、实验数据和变化趋势等),只考虑系统变量随时间的发展变化规律,对其未来作出预测。主要包括移动平均法、指数平滑法、趋势外推法以及博克斯—詹金斯(Box-Jenkins)法等。

③因果关系预测方法。系统变量之间存在着某种前因后果的关系,找出影响某种结果的一个或几个因素,建立起它们之间的数学模型,然后可以根据自变量的变化预测结果变量的变化。因果关系模型中的因变量与自变量在时间上是同步的,即因变量的预测值要由相应的自变量的值来旁推。因果关系预测方法主要有线性回归法、马尔可夫法、状态空间预测法、计量经济预测法以及系统动力学仿真方法等。

(2)预测分析一般步骤

①明确预测目标。要在整个决策问题研究的总目标指导下,确定预测对象及具体的要求,包括预测指标、预测期限、可能选用的预测方法以及需要的基本资料和数据。

②收集、整理资料和数据。收集有关的历史资料、统计数据、试验数据等,并进行整理分析,进行调查、访问,以取得第一手数据资料。

③建立预测模型。根据科学理论指导以及所选择的预测方法,用各种有关变量来真实表达预测对象的关系,从而建立起预测所需的数学模型。

④模型参数估计。按照各种模型的性质和可能的样本数据,对模型中的参数进行估计,最终识别和确认所选用的模型形式和结构。

⑤模型检验。检验包括对模型的合理性及有效性的验证,一是对有关假设的检验,如对线性关系的假设、变量结构(变量选取)以及独立性的假设等必须进行统计检验,以保证理论、方法的正确性;二是对模型精确度即预测误差的检验,如误差区间、标准离差等的检验。经检验一旦发现模型不合理,就必须对模型加以修正。

⑥预测实施与结果分析。运用通过检验的预测模型,使用有关数据进行预测,并对预测结果进一步进行有关理论、经验方面的分析。必要时还可以对不同方法模型的预测结果加以分析对比,以作出更加可信的判断。

(3)定性预测方法——德尔斐法

定性预测是以人的逻辑判断为主,并根据由各种途径得到的意见、信息和有关资料,综合分析当前的政治、经济、科技等形势以及预测对象的内在联系,以判断事件发展的前景,并尽量把这种判断转化为可计量的预测。定性预测法一般适用于缺乏历史统计数据的预测对象。德尔斐法是最典型的一种定性预测方法。

德尔斐法是专家会议调查法的一种发展。它以匿名方式通过几轮函询，征求专家的意见。预测领导小组对每一轮意见都进行汇总整理，作为参考资料再发给每位专家，供他们分析判断，提出新的论证。如此多次反复，专家意见趋于一致，结论的可靠性越来越大。

德尔斐法有如下三个特点：①为克服专家会议易受心理因素影响的缺点，德尔斐法采用匿名形式；②德尔斐法不同于民意测验，一般要经过 3～4 轮；③为了定量评价预测结果，德尔斐法采用统计方法对结果进行处理。

例 10.1 某公司研制出一种新兴产品，现在市场上还没有相似产品出现，因此没有历史数据可以获得。公司需要对可能的销售量做出预测，以决定产量。于是该公司成立专家小组，并聘请业务经理、市场专家和销售人员等 8 位专家，预测全年可能的销售量。8 位专家提出个人判断，经过三次反馈得到结果如下表所示。

表 10-4 共三轮 8 位专家打分结果

专家编号	第一次判断			第二次判断			第三次判断		
	最低销售量	最可能售量	最高销售量	最低销售量	最可能售量	最高销售量	最低销售量	最可能售量	最高销售量
1	500	750	900	600	750	900	550	750	900
2	200	450	600	300	500	650	400	500	650
3	400	600	800	500	700	800	500	700	800
4	750	900	1500	600	750	1500	500	600	1250
5	100	200	350	220	400	500	300	500	600
6	300	500	750	300	500	750	300	600	750
7	250	300	400	250	400	500	400	500	600
8	260	300	500	350	400	600	370	410	610
平均数	345	500	725	390	550	775	415	570	770

在预测时，最终一次判断是综合前几次的反馈作出的，因此在预测时一般以最后一次判断为主。

平均值预测：按照 8 位专家第三次判断的平均值计算，则预测新产品的平均销售量为：

(415＋570＋770)/3＝585(千件)

加权平均预测：将最可能销售量、最低销售量和最高销售量分别按 0.50、0.20 和 0.30 的概率加权平均，则预测平均销售量为：

0.2×415+0.5×570+0.3×770=599(千件)

中位数预测:用中位数计算,可将第三次判断按预测值高低排列如下:

最低销售量:300、300、370、400、400、500、500、550

最可能销售量:410、500、500、500、600、600、700、750

最高销售量:600、600、610、650、750、800、900、1250

最低销售量的中位数为第四项与第五项的平均值,即(400+400)/2=400

最可能销售量的中位数为(500+600)/2=550

最高销售量的中位数为(650+750)/2=700

将最可能销售量、最低销售量和最高销售量分别按 0.50、0.20 和 0.30 的概率加权平均,则预测平均销售量为:

0.2×400+0.5×550+0.3×700=565(千件)

(4)回归分析法

我们仅简单介绍最基本的一元线性回归模型。

假设 y 与 x 可能存在线性相关 $y=a+bx$,但是由于实验误差或其他原因,使得 y 与 x 变成近似线性关系 $y=a+bx+\varepsilon$,其中未知参数 a、b 是需要设法估算的,ε 为误差项。

设成对观测样本数据为$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$,则利用最小二乘法可估计参数 a,b:

$$\begin{cases}\hat{b}=\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]/\sum_{i=1}^{n}(x_i-\bar{x})^2,\\ \hat{a}=\bar{y}-\hat{b}\bar{x},\bar{x}=\sum_{i=1}^{n}x_i/n,\bar{y}=\sum_{i=1}^{n}y_i/n\end{cases}$$

得一元经验线性回归方程:$y=\hat{a}+\hat{b}x$。

例 10.2 已知身高与体重的资料如下表:

表 10-5 身高与体重

y 身高(cm)	155	160	165	167	170	175	180	182
x 体重(kg)	50	52	57	56	60	65	62	70

试计算:(1)拟合 y 关于 x 的线性回归方程;(2)当体重为 75 kg 时,预测其身高。

解:代入公式计算得 $\hat{b}=0.0134$,$\hat{a}=0.898$;经验回归方程 $\hat{y}=0.898+0.0134x$。也就是说身高每增加 1 cm,体重平均增加 0.0134 kg。

将 $x=75$ kg 代入经验回归方程得

$$\hat{y}=0.898+0.0134\times75=190.3(\text{cm})。$$

10.5.3 决策

(1)决策基本知识

为了实现特定的目标，根据客观的可能性，在占有一定信息和经验的基础上，借助一定的工具、技巧和方法，对影响未来目标实现的诸因素进行准确的计算和判断选优后，对未来行动作出决定。

决策的三个基本特征：未来性、选择性和实践性。

组成决策系统的四个基本要素：决策主体、决策目标、决策对象和决策环境。

统计决策的概念有广义和狭义之分。凡是使用统计方法进行决策的决策方法称为广义的统计决策；狭义的统计决策是指不确定情况下的决策。

不确定情况下的决策需要具备的四个条件：决策人要求达到的一定目标；存在两个或两个以上可供选择的方案；存在着不以决策人主观意志为转移的客观状态，即自然状态；在不同情况下采取不同方案所产生的结果是可以计量的。

从不同的角度决策分类有：

①按决策问题所处的条件分为确定型决策、不确定型决策和对抗型决策；

②按问题的性质分为程序化决策和非程序化决策；

③按决策涉及的范围分为总体决策和局部决策；

④按决策过程是否运用数学模型来辅助决策分为定性决策和定量决策；

⑤按决策目标的数量分为单目标决策和多目标决策；

⑥按决策的整体构成分为单阶段决策和多阶段决策。

(2)例子

例 10.3 某厂为适应市场的需要，准备扩大生产能力，有两种方案可供选择：第一方案是建大厂；第二方案是先建小厂，后考虑扩建。如建大厂，需投资 700 万元，在市场销路好时，每年收益 210 万元，销路差时，每年亏损 40 万元。在第二方案中，先建小厂，如销路好，3 年后进行扩建。建小厂的投资为 300 万元，在市场销路好时，每年收益 90 万元，销路差时，每年收益 60 万元，如果 3 年后扩建，扩建投资为 400 万元，收益情况同第一方案一致。未来市场销路好的概率为 0.7，销路差的概率为 0.3；如果前 3 年销路好，则后 7 年销路好的概率为 0.9，销路差的概率为 0.1。无论选用何种方案，使用期均为 10 年，试做决策分析。

解：这是一个多阶段的决策问题，考虑采用期望收益最大为标准选择最优方案。

第一步，画出决策树图。

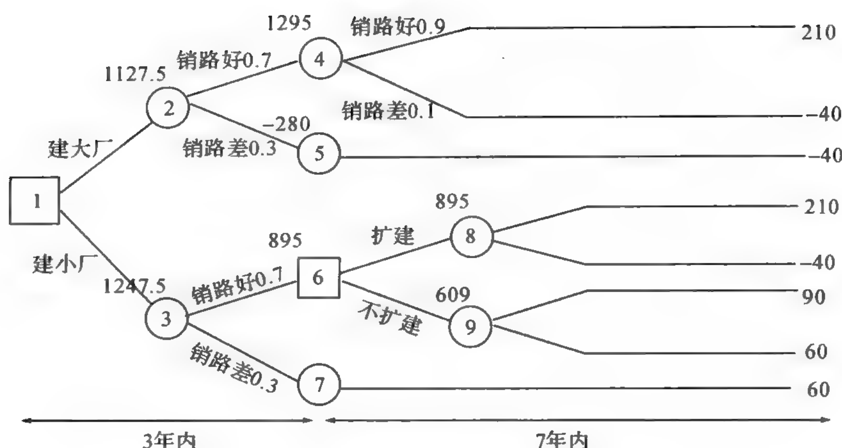


图 10-2　决策树

第二步，从右向左计算各点的期望收益值。

点 4：210×0.9×7－40×0.1×7＝1295(万元)

点 5：－40×7＝－280(万元)

点 2：1295×0.7＋210×0.7×3－280×0.3－40×0.3×3＝1227.5(万元)

点 8：210×0.9×7－40×0.1×7－400＝895(万元)

点 9：90×0.9×7＋60×0.1×7＝609(万元)

点 6 是个决策点，比较点 8 和点 9 的期望收益，选择扩建。

点 6：895(万元)

点 7：60×7＝420(万元)

点 3：895×0.7＋210×0.7×3＋420×0.3＋60×0.3×3＝1247.5(万元)

第三步，进行决策。

比较点 2 和点 3 的期望收益，点 3 期望收益值较大，可见，最优方案是先建小厂，如果销路好，3 年以后再进行扩建。

思考题

1. 下载一款股票交易软件(如同花顺)，从交易软件中下载上证综合指数交易历史数据，参考 10.2 节方法，研究最近 10 年上证综指波段规律。

2. 设法下载美国道琼斯指数、香港恒生指数历史交易数据，研究它们的波段特征，并与上证综指进行比较，重点比较三者波段划分、波段振幅的异同性。

3. 研究股指期货主力持仓与上证综指开盘价的关联度。

4. 针对例 10.3 解题过程，抽象提炼成一种一般化的决策过程。

第11章 数学建模竞赛

国家教育部高教司、各省市教育主管部门与其他行政部门及各种学会(协会)经常组织各类(学科)竞赛,如:大学生数学建模竞赛、大学生电子设计竞赛、大学生机械设计竞赛、大学生结构设计竞赛、大学生程序设计竞赛、大学生机器人竞赛、大学生创业计划竞赛、大学生电子商务竞赛、大学生多媒体作品竞赛、大学生财会信息化竞赛、大学生演讲比赛、大学生英语竞赛、大学生课外学术科技作品竞赛、大学生广告艺术竞赛和大学生智能汽车竞赛等。其中数学建模竞赛是学生参赛人数最多的竞赛之一。

各教育部门、高等院校组织各种竞赛的目的,是为了加强第二课堂活动,构建大学生实践创新平台,营造和丰富校园科技文化氛围,多方面、多途径拓展大学生的综合知识,培养大学生自主创新意识、创新思维、实践动手能力和团队合作精神,提高竞赛组织管理和参赛水平,造就21世纪"知识、能力、素质"三者协调发展的具有创新创业精神的人才。

11.1 数学建模竞赛

11.1.1 美国数学建模竞赛

美国大学生数学建模竞赛(MCM/ICM)是一项国际级的竞赛项目,为现今各类数学建模竞赛之鼻祖。MCM/ICM 是 Mathematical Contest in Modeling 和 Interdisciplinary Contest in Modeling 的缩写,即"数学建模竞赛"和"交叉学科建模竞赛"。1985年,在美国科学基金会的资助下,创办了一个名为"数学建模竞赛"(Mathematical Competition in Modeling 后改名 Mathematical Contest in Modeling,简称 MCM),是一年一度的大学水平的竞赛。ICM 始于2000年,它们都由 COMAP(The Consortium for Mathematics and Its Application,美国数学及其应用联合会)主办,并得到了美国国家自然科学基金会、阿尔佛雷德基金会、艾克森美孚基金会、美国卡内基公司、杰拉尔丁道奇基金会、美国国家安全局、美国教育部等的资助,同时受到美国统计学会、康科德财团、弗洛伊登萨尔研究所、运筹学和管理科学学会、数学教育

国际会议、数学应用教学国际会议、美国数学会、数学/自然科学教育委员会、美国工业与应用数学会等协助。官方网址：http://www.comap.com/index.html。

MCM/ICM 着重强调研究问题、解决方案的原创性、团队合作、交流以及结果的合理性。竞赛要求以三人组成一队的形式参赛，在 4 天(96 小时)(早年为 3 天，即 72 小时)时间内，就指定的问题完成从建立模型、求解、验证到论文撰写的全部工作。它是一种彻底公开的竞赛，每年只有 3 个来自不受限制的任何领域的实际问题，学生任选 1 题，完成该实际问题的数学建模的全过程，并就问题的重述、简化和假设及其合理性的论述、数学模型的建立和求解(及软件)、检验和改进、模型的优缺点及其可能的应用范围的自我评述等内容写出论文。

由专家组成的评阅组进行评阅，评出优秀论文，并给予某种奖励。它只有唯一的禁律，就是在竞赛期间不得与队外任何人(包括指导教师)讨论赛题，但可以利用任何图书资料、互联网上的资料、任何类型的计算机和软件等，为充分发挥参赛学生的创造性提供了广阔的空间。第一届 MCM 时，美国 70 所大学 90 个队参加，竞赛每年都吸引大量著名高校参赛，2012 年全球共有来自美国、加拿大、中国、印度、德国、英国、新加坡、墨西哥、爱尔兰、西班牙等 18 个国家的 5026 支代表队参加此项赛事。从某种意义上讲，它已经成为一种国际性的竞赛，影响极其广泛。

美国大学生数学建模竞赛共设置五个奖项，分别为 Outstanding Winner，Finalist，Meritorious Winner，Honorable Mentions，Successful Participant。在国内，约定俗成地将这五个奖项分别对应为特等奖、特等候选奖、一等奖、二等奖、三等奖。

美国大学生数学建模竞赛每年的比赛时间一般定在二月初，需要通过官方网站报名，而且需要有固定的指导教师。一般各大高校均会组织感兴趣的同学进行赛前培训以及报名、交费等事宜。比赛时间：2011 年 2 月 11—15 日，2012 年 2 月 10—14 日，2013 年 2 月 1—5 日。

11.1.2 中国数学建模竞赛

1989 年 2 月 24—26 日，中国大学生(北京大学、清华大学、北京理工大学共 4 个队)首次参加美国大学生数学建模竞赛，自此每年我国都有学生参加美国数学建模竞赛。1990 年 12 月 7—9 日，上海市举办大学生(数学类)数学模型竞赛，这是我国省、市级首次举办数学建模竞赛。1992 年 11 月 27—29 日，部分城市大学生数学模型联赛举行，这是全国性的首届竞赛，10 省(市)74 所

院校的314队参加。

1993年12月16日,国家教委高教司发出教高司〔1993〕178号文件《关于进行"电子设计"等四项竞赛筹备工作的通知》,其中包括数学建模竞赛。从此全国大学生数学建模竞赛成为全国性大学生学科竞赛活动,每年9月的某个周五8:00至下周一8:00,连续72小时。2012年,来自全国33个省、市、自治区(包括香港和澳门地区)及新加坡、美国的1284所院校21219个队(其中本科组17741队、专科组3478队)63600多名大学生报名参加本项竞赛。官方网址:http://www.m cm.edu.cn/index_cn.html。

全国大学生数学建模竞赛的竞赛宗旨:创新意识,团队精神,重在参与,公平竞争。

全国大学生数学建模竞赛的指导原则:扩大受益面,保证公平性,推动教学改革,提高竞赛质量,扩大国际交流,促进科学研究。

11.1.3 全国大学生数学建模竞赛章程(2008年修订)

第一条　总则

全国大学生数学建模竞赛(以下简称竞赛)是教育部高等教育司和中国工业与应用数学学会共同主办的面向全国大学生的群众性科技活动,目的在于激励学生学习数学的积极性,提高学生建立数学模型和运用计算机技术解决实际问题的综合能力,鼓励广大学生踊跃参加课外科技活动,开拓知识面,培养创造精神及合作意识,推动大学数学教学体系、教学内容和方法的改革。

第二条　竞赛内容

竞赛题目一般来源于工程技术和管理科学等方面经过适当简化加工的实际问题,不要求参赛者预先掌握深入的专门知识,只需要学过高等学校的数学课程。题目有较大的灵活性供参赛者发挥其创造能力。参赛者应根据题目要求,完成一篇包括模型的假设、建立和求解、计算方法的设计和计算机实现、结果的分析和检验、模型的改进等方面的论文(即答卷)。竞赛评奖以假设的合理性、建模的创造性、结果的正确性和文字表述的清晰程度为主要标准。

第三条　竞赛形式、规则和纪律

1. 全国统一竞赛题目,采取通讯竞赛方式,以相对集中的形式进行。

2. 竞赛每年举办一次,一般在某个周末前后的三天内举行。

3. 大学生以队为单位参赛,每队3人(须属于同一所学校),专业不限。竞赛分本科、专科两组进行,本科生参加本科组竞赛,专科生参加专科组竞赛(也可参加本科组竞赛),研究生不得参加。每队可设一名指导教师(或教师

组)，从事赛前辅导和参赛的组织工作，但在竞赛期间必须回避参赛队员，不得进行指导或参与讨论，否则按违反纪律处理。

4. 竞赛期间参赛队员可以使用各种图书资料、计算机和软件，在国际互联网上浏览，但不得与队外任何人(包括在网上)讨论。

5. 竞赛开始后，赛题将公布在指定的网址供参赛队下载，参赛队在规定时间内完成答卷，并准时交卷。

6. 参赛院校应责成有关职能部门负责竞赛的组织和纪律监督工作，保证本校竞赛的规范性和公正性。

余略。

11.1.4 全国研究生数学建模竞赛

全国研究生数学建模竞赛是在全国范围内举行的、面向研究生群体的赛事活动，该赛事得到了全国各研究生培养单位的大力支持和各地研究生的热烈响应。2004 年举办第 1 届研究生数学建模竞赛，2006 年被列为教育部研究生教育创新计划项目之一。近几年，随着这项赛事的迅速发展和广泛宣传，参赛学校及参赛队伍的规模越来越大，参赛者中不乏名校博士生及中科院的研究生，使这项赛事在全国高校研究生中的影响力越来越大，在广大研究生中也打下了良好的群众基础。官方网址：http://gmcm.seu.edu.cn/s/274/t/1419/main.htm。

东南大学现为该项赛事的组委会主任委员单位，主要负责每一届赛事的组织工作。赛事的承办工作由全国研究生数学建模竞赛组委会成员单位轮流负责。

研究生数学建模竞赛活动成立有学术委员会和组织委员会，分别负责活动的组织和学术工作。整个活动分五个阶段。

11.2 数学建模竞赛题目

数学建模竞赛题目多数来自实际问题，考虑到竞赛时间的限制，竞赛题已是实际问题的简化。本节只给出 2012 年美国赛和中国赛的全部题目，目的是让读者了解一次竞赛的题目搭配情况。详细的内容可浏览竞赛官方网站，历年竞赛题目、要求、结果等都可以查到。研究生数学建模竞赛题目更接近科研课题，题目都很长，略。

美国赛没有具体明确地指出评审原则，但假设合理性、建模的创造性和文字表述的清晰程度非常重要，在其特等奖的论文中某些结果相差 50%、

100%，甚至相差更大都有可能出现。关键要做到自圆其说。

中国赛则条条框框更多些，这样容易评阅。为此我们特地在全国赛题目后面附上由竞赛委员会给出的评阅要点以供参考。

11.2.1 美国数学建模竞赛题目

美国竞赛每年共有3道题目A、B、C，其中题A与题B属于MCM，一般题目较短；题C属于ICM，通常题目较长且原始数据特多。

(1)2012美赛A题

一棵树的叶子

“一棵树的叶子有多重?”怎么能估计树的叶子(或者树的任何其他部分)的实际重量?怎样对叶子进行分类?建立一个数学模型来对叶子进行描述和分类。模型要考虑和回答下面的问题。

为什么叶子具有各种形状?

叶子之间是要将相互重叠的部分最小化，以便可以最大限度地接触到阳光吗?树叶的分布以及树干和枝杈的体积影响叶子的形状吗?

就轮廓来讲，叶形(一般特征)是和树的轮廓以及分枝结构有关吗?

你将如何估计一棵树的叶子质量?叶子的质量和树的尺寸特征(包括和外形轮廓有关的高度、质量、体积)有联系吗?

除了你的一页摘要以外，给科学杂志的编辑写一封信，阐述你的主要发现。

(2)2012美赛B题

沿着“大长河”露营

游客在“大长河”(225英里)可以享受到秀丽的风光和令人兴奋的白色湍流。这条河对于背包客来说是进不去的，因此畅游这条长河的唯一办法就是在这条河上露营几天。这次旅行从开始的下水点到最终结束点，共225英里，且是顺流而下的。乘客可以选择平均4英里/小时的、以桨作为动力的橡胶筏或者平均8英里/小时的机动帆船旅行。整个旅行从开始到结束会经历6至18个夜晚。负责管理这条河的政府机构希望到这里的每一次旅行都能够享受到野外经历，以最少接触到在河上的其他船只。目前，每年在六个月期间(一年的其余时间的天气对于河流旅行来说太冷)，共有X次旅行，有Y处露营地，露营地均匀地分布在整个河道。由于漂流受欢迎程度的上升，公园管理者已经被要求允许更多的旅行次数。所以他们想确定怎样可能安排一个最优的混合的旅行方案，不同的时间(单位为夜)和推动方式(马达或桨)，最大限度地利用露营地。换句话说，在长河的漂流季，能有多少游客加入乘船

旅行？河流的管理者现在雇佣你，为他们提出最佳排程方式和河流承载能力的建议，记住两个露营者不能在同一时间内占据同一个露营地。除了你的一页摘要，准备一页备忘录，向河流的管理者描述你的主要发现。

(3)2012 美赛 C 题

犯罪克星

你所在的组织银河犯罪建模中心(ICM)正在调查一个实施犯罪行为的阴谋。调查人员现在非常有信心，他们已经知道策划阴谋的一些成员，但是他们希望在逮捕嫌疑人之前确定其他的犯罪成员和组织的领导人。所有的嫌疑人和可能涉嫌的同谋都受雇于同一家公司，在一个大的综合办公室里工作。该公司发展迅速，正在开发和销售以自己的名字命名的计算机软件，该软件是为银行和信用卡公司服务的。ICM 最近从公司的一组员工(有 82 人)那里获得了一些消息，他们认为这将帮助他们找到最有可能的未知身份的同谋者和组织领导人。由于公司中的所有员工都知晓该消息，所以一些消息的传播者(有可能很多)并没有卷入阴谋。事实上，他们可以确定有一些人没有卷入阴谋。建模工作的目标是确定在综合办公室里面的人谁最有可能是同谋者。一个优先级列表是最理想的，ICM 可以按照优先级调查、监视或者审问最有可能的嫌疑人。一个判别是否为同谋人的分界线也非常有用，可以用它来对各组人进行分类。对于检方来讲，如果模型能够识别出阴谋策划的领导人也是非常有帮助的。在你的犯罪建模团队获得当前案件的数据之前，你的上司给了你们下面的一些场景(被称作调查 EZ)，这些场景是几年前她在其他城市工作时遇到的。尽管她对自己在 EZ 案件上的工作非常自豪，她仍然谦虚地说那是一个小的、简单的案例，但它可以帮助你了解你的任务。她的数据如下。

她考虑为同谋者的十个人分别为：Anne# (安妮)，Bob(鲍勃)，Carol(卡罗尔)，Dave* (戴夫)，Ellen(艾伦)，Fred(弗雷德)，George* (乔治)，Harry(哈里)，Inez(伊内兹)，Jaye# (杰伊)(* 表示事先已知是同谋者，# 表示事先已知为非同谋者)。下面是 28 条消息的列表，这些消息是在她的案件中获得的，每条消息后面有一个标号，这个标号反映了她对消息主题的分析。

安妮对鲍勃说：为什么你今天迟到了？(1)

鲍勃对卡罗尔说：这该死的安妮总是看着我。我没有迟到。(1)

卡罗尔对戴夫说：对于鲍勃的迟到，安妮和鲍勃有争执。(1)

戴夫对艾伦说：我需要今天早晨看见你。什么时间你能来？把预算文件带来。(2)

戴夫对弗雷德说：今天我随时都可以来见你。如果时间合适就告诉我。

我应该把预算文件带上吗？(2)

戴夫对乔治说：我之后要见你……有很多话要说。我希望其他人做好准备。重要的是要得到这个权利。(3)

哈里对乔治说：你似乎在强调。这是怎么回事？我们的预算很优秀。(2)(4)

伊内兹对乔治说：我今天真的感觉很累。你感觉怎么样？(5)

杰伊对伊内兹说：今天没有太多事可做。去吃午餐怎么样？(5)

伊内兹对杰伊说：好想法，但是我筋疲力尽了，不能做午餐了，对不起！(5)

乔治对戴夫说：谈话时间，现在！(3)

杰伊对安妮说：你今天能去吃午餐吗？(5)

戴夫对乔治说：我不能。我要在回家的路上去看弗雷德。(3)

乔治对戴夫说：那之后到这。(3)

安妮对卡罗尔说：谁应该去看看鲍勃？他正在消磨时间。(1)

卡罗尔对安妮说：别理他。他和乔治、戴夫相处得很好。(1)

乔治对戴夫说：这是非常重要的。该死的弗雷德。艾伦如何？(3)

艾伦对乔治说：你和戴夫谈话了？(3)

乔治对艾伦说：还没有。你呢？(3)

鲍勃对安妮说：我没有迟到。我要让你知道，午餐时间我也在工作。(1)

鲍勃对戴夫说：告诉他们我没有迟到。你知道的。(1)

艾伦对卡罗尔说：去找安妮，弄清楚下周预算会议的具体日程，并且帮助我让乔治冷静一下。(2)

哈里对戴夫说：你没有注意到今天乔治又压力很大？(4)

戴夫对乔治说：该死的哈里认为你压力很大。不要让他担心或者别让他察觉。(4)

乔治对哈里说：仅仅是因为工作太晚了，家里也出了点问题。我很好。(4)

艾伦对哈里说：如果我错过了今天的会议，一切还会都好吗？弗雷德将出席会议，他知道预算且比我做得好。(2)

哈里对弗雷德说：我认为明年的财政预算案使一些人压力很大。也许我们应该花些时间来让人们放心。(2)(4)

弗雷德对哈里说：我认为我们的预算是非常健康的。我没有看到任何压力。(2)

你的上司指出，她分配并编码了仅仅 5 种不同的消息主题：①鲍勃的迟

到；②预算；③重要但未知的问题，被认为是阴谋的一部分；④乔治的压力；⑤午餐和其他社会问题。正如所看到的消息编码，一些消息因为其内容和两个主题联系在了一起。

你的上司分析情况采用的方法是一个网络，它显示了消息的通讯连接情况和消息的类型。下图是一个消息网络模型，网络图上注明了消息类型的代码。

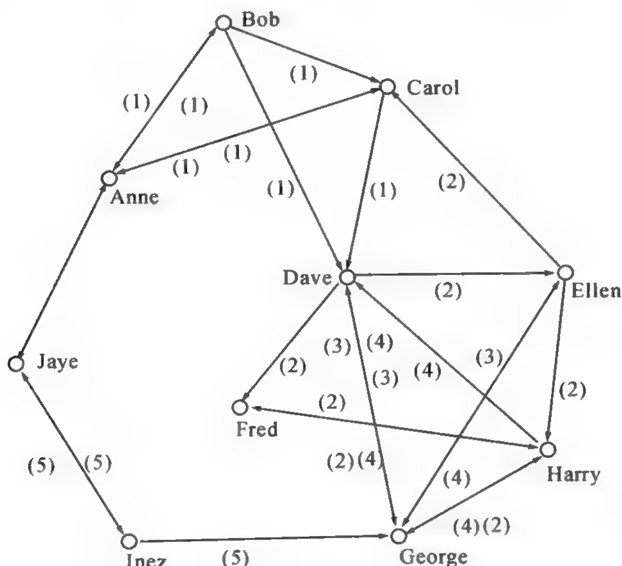


图 11-1　EZ 案件的消息网络

你的上司指出，除了已知的同谋乔治和戴夫，根据她的分析，艾伦和卡罗尔也因为同谋而被起诉，后来鲍勃自己认罪，被判处减刑。但对卡罗尔的起诉后来被撤销了。你的上司仍然坚信伊内兹参与其中，但对于她的诉讼始终没能成立。你的上司给你们团队建议，一定要明确人群中有罪的一部分人，像伊内兹这样的人不能漏网，像卡罗尔这样的人也不能被错误地起诉，并且ICM 得到证据，像鲍勃这样的人就没有机会获得减刑。

目前情况下，你的上司已经得到了一个网络形式的数据库，它有着相同的结构，但是在规模上稍大一些。有一些迹象表明，一个阴谋正在从公司挪用资金，并且使用网络欺诈窃取与公司做生意的人的信用卡内的资金。她给你示范了一个小例子，在 EZ 情况下只有 10 个人(节点)，27 条边(信息)，5 个主题，一个可疑的(阴谋的)主题，2 个已知的同谋者，还有 2 个已知的非同谋者。目前，这个新的情况，有 83 个节点，400 条边(其中一些包含不止一个主

题)，超过21000个字符的信息传输，15个主题(3个被视为是可疑的)，7个是已知的同谋者，还有8个已知的非同谋者，数据在给出的附件 Topics. xls, Messages. xls 和 Names. xls 中，Names. xls 包含办公室员工的姓名，和节点的数目一样。Topics. xls 包含了15个主题的代码和简短描述。由于安全和隐私的问题，你的团队将不能得到所有信息流的副本。Messages. xls 提供链接节点的用来传递信息的边，信息中包含数字代码。一些信息包含了三个主题。为了可视化信息流动，对于人和信息传播的网络模型如图11-2所示。图上没有像图11-1一样标注消息的主题。这些主题的编号在文件 Messages. xls中给出，主题描述在 Topics. xls 中给出。

要求一：目前，已知 Jean, Alex, Elsie, Paul, Ulf, Yao 和 Harvey 是同谋者，还知道 Darlene, Tran, Jia, Ellin, Gard, Chris, Paige 和 Este 不是同谋者。3个已知的可疑信息主题为7、11和13，更多的主题细节请见附件 Topics. xls。根据83个节点为阴谋的可能性的大小，建立一个模型和算法对可能性大小进行排序，并说明你的模型及流程。Jerome, Delores 和 Gretchen 是公司的高级管理人员，如果能知道他们中的任何一个是否参与了这个阴谋将是十分有用的。

要求二：如果得到新的消息，主题一和阴谋有关，且 Chris 是同谋者之一，请问要求一中的排序会是什么样子的?

要求三：和这个消息传输类似的一种强大的用来获得和理解文本信息的技术被称为语义网消息传输分析。作为一个人工智能和计算语言学的方法，它为知识推理和语言提供了一个结构和过程。另一种在自然语言处理能力方面的计算语言学叫做文本分析。在我们的犯罪现场破坏条件下，解释语义和文本内容的分析及消息传输的背景(如果你能获得原始信息)能使您的团队开发更好的关于办公室人员的模型和归类。你是否在文件 Topics. xls 中有使用这些功能来进行主题描述，从而提升你的模型?

要求四：你完成的报告最终将送给检察官，所以报告中必须详细地、清晰地陈述你的假设和方法论。但是报告不能超过20页。你可以在分开的文件中包含你的程序作为附录，附录不算在页数内，但这些附录不是必要的。你的上司希望 ICM 在解决白领高技术的阴谋犯罪方面是世界上最好的。希望你提供的方法将有助于解决世界各地重大案件，尤其是那些拥有非常大消息传输的数据库(成千上万的人，成千上万的信息，可能数百万字)。她特别要求你在报告中要包含可以帮助你的信息模型和建议的讨论，包含更深入的网络、语义和文本内容分析的内容。作为你向她报告的一部分，说明你使用的网络建模技术，和你为什么使用它以及如何使用它在任何类型的网络数据库

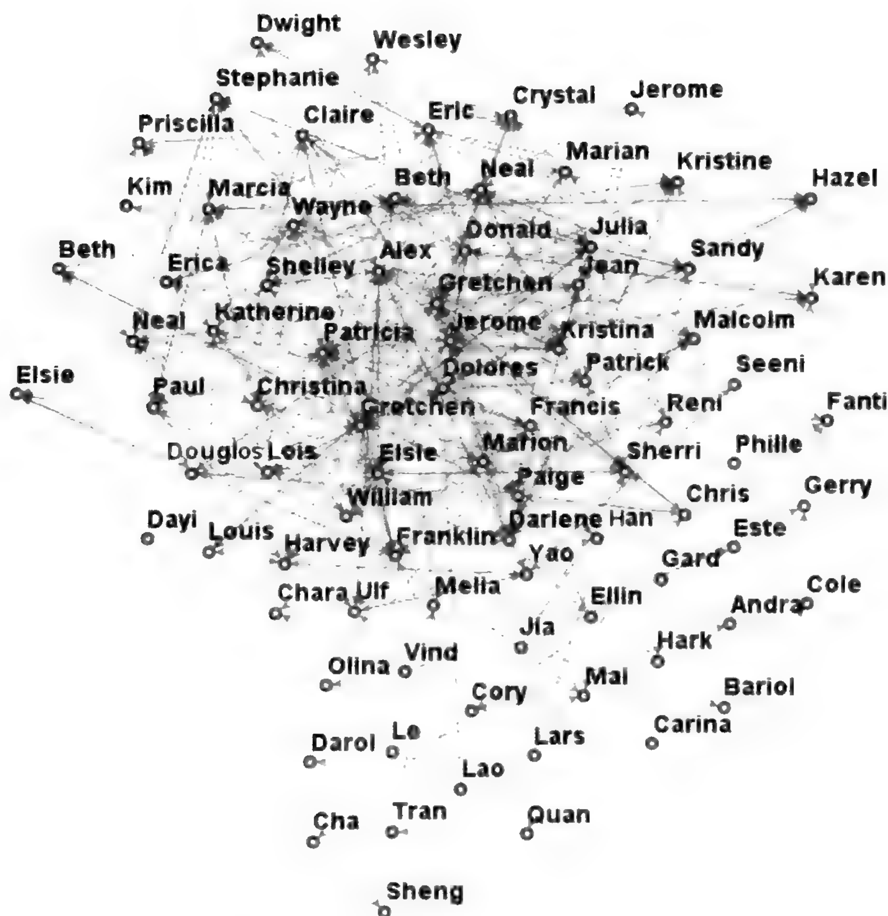


图 11-2 可视化的网络模型，包含 83 个人(节点)和 400 条他们之间的信息(边)

中用来识别、排序和分类相似节点，而不仅仅是犯罪阴谋和信息数据。例如，在得到节点感染概率和部分已经确认感染节点的各种图像或化学数据的生物网络中，你的方法能否找到感染或患病的细胞的位置？

你提交的 ICM 论文应该包含一页摘要和不超过 20 页的解决方案，总计不超过 21 页。

11.2.2 全国大学生数学建模竞赛题目

(1)全国赛 A 题

葡萄酒的评价

确定葡萄酒质量时一般是通过聘请一批有资质的评酒员进行品评。每个评酒员在对葡萄酒进行品尝后对其分类指标打分,然后求和得到其总分,从而确定葡萄酒的质量。酿酒葡萄的好坏与所酿葡萄酒的质量有直接的关系,葡萄酒和酿酒葡萄检测的理化指标会在一定程度上反映葡萄酒和葡萄的质量。附件 1 给出了某一年份一些葡萄酒的评价结果,附件 2 和附件 3 分别给出了该年份这些葡萄酒的和酿酒葡萄的成分数据。请尝试建立数学模型讨论下列问题。

1. 分析附件 1 中两组评酒员的评价结果有无显著性差异,哪一组结果更可信?

2. 根据酿酒葡萄的理化指标和葡萄酒的质量对这些酿酒葡萄进行分级。

3. 分析酿酒葡萄与葡萄酒的理化指标之间的联系。

4. 分析酿酒葡萄和葡萄酒的理化指标对葡萄酒质量的影响,并论证能否用葡萄和葡萄酒的理化指标来评价葡萄酒的质量?

附件 1:葡萄酒品尝评分表(含 4 个表格)。(略)

附件 2:葡萄和葡萄酒的理化指标(含 2 个表格)。(略)

附件 3:葡萄和葡萄酒的芳香物质(含 4 个表格)。(略)

[评阅要点]

本题目希望学生利用数学模型和附件 1—3 中的数据对评酒员的品评结果给出分析,对酿酒葡萄的质量给出评价,并探讨葡萄和葡萄酒的理化指标与酒的质量的关系。

问题 1:附件 1 中给出的是评酒员对 27 种红葡萄酒和 28 种白葡萄酒的两组品评结果。这两组评酒员各不相同,两组中的每个酒样都取自相同葡萄酒厂家的同一批次的产品。要求学生给出判断这两组评价结果好坏的原理、模型和方法,给出具体的结果,并对结果进行说明。好的品评结果应该是对同一酒样评价时这些评酒员之间的差距小、且这些酒样之间的区分度明确(注:一些学生的模型和方法仅考虑评酒员的打分差距)。参考:红酒中样品 23 是好酒,样品 12 是较差的酒。

问题 2:给出根据酿酒葡萄的理化指标和葡萄酒的质量对这些酿酒葡萄进行分级的原则、模型、算法和结果。确定酿酒葡萄质量好坏的主要依据是问题 1 中评酒员对酒的质量的评价结果,根据这个评价结果和酿酒葡萄的各

种理化指标给出确定葡萄质量的模型，由此给出这些酿酒葡萄的分级结果。参考：分级结果中好的红葡萄应包含样品 23，差的应该包含样品 12。

问题 3：给出分析酿酒葡萄与葡萄酒的成分之间关系的原理、模型和方法，得到葡萄酒的理化指标是否与葡萄的理化指标相关的结论，相关时给出具体的依赖关系。求解时最好先对葡萄的理化指标（包括芳香物质）进行分类和筛选，然后进行评价。注：仅把葡萄的全部理化指标进行简单回归是不够完整的。

问题 4：建立模型分析酿酒葡萄和葡萄酒的理化指标与葡萄酒质量之间的关系，在模型的基础上给出具体结论，并对结论给出详细的分析说明。注：评价葡萄酒质量时不一定需要包含所有的理化指标，但根据经验知道花色苷、总酚和单宁是红葡萄酒的重要指标。

附注：学生答卷中应该说明对缺失数据和异常数据的处理方式。

(2)全国赛 B 题

太阳能小屋的设计

在设计太阳能小屋时，需在建筑物外表面（屋顶及外墙）铺设光伏电池，光伏电池组件所产生的直流电需要经过逆变器转换成 220 V 交流电才能供家庭使用，并将剩余电量输入电网。不同种类的光伏电池每峰瓦的价格差别很大，且每峰瓦的实际发电效率或发电量还受诸多因素的影响，如太阳辐射强度、光线入射角、环境、建筑物所处的地理纬度、地区的气候与气象条件、安装部位及方式（贴附或架空）等。因此，在太阳能小屋的设计中，研究光伏电池在小屋外表面的优化铺设是很重要的问题。

附件 1—7 提供了相关信息。请参考附件提供的数据，对下列三个问题分别给出小屋外表面光伏电池的铺设方案，使小屋的全年太阳能光伏发电总量尽可能大，而单位发电量的费用尽可能小，并计算出小屋光伏电池 35 年寿命期内的发电总量、经济效益（当前民用电价按 0.5 元/(kW · h)计算）及投资的回收年限。

在求解每个问题时，都要求配有图示，给出小屋各外表面电池组件铺设分组阵列图形及组件连接方式（串、并联）示意图，也要给出电池组件分组阵列容量及选配逆变器规格列表。

在同一表面采用两种或两种以上类型的光伏电池组件时，同一型号的电池板可串联，而不同型号的电池板不可串联。在不同表面上，即使是相同型号的电池也不能进行串、并联连接。应注意分组连接方式及逆变器的选配。

问题 1：请根据山西省大同市的气象数据，仅考虑贴附安装方式，选定光伏电池组件，对小屋（见附件 2）的部分外表面进行铺设，并根据电池组件分组

数量和容量，选配相应的逆变器的容量和数量。

问题 2：电池板的朝向与倾角均会影响到光伏电池的工作效率，请选择架空方式安装光伏电池，重新考虑问题 1。

问题 3：根据附件 7 给出的小屋建筑要求，请为大同市重新设计一个小屋，要求画出小屋的外形图，并对所设计小屋的外表面优化铺设光伏电池，给出铺设及分组连接方式，选配逆变器，计算相应结果。

附件 1：光伏电池组件的分组及逆变器选择的要求。（略）

附件 2：给定小屋的外观尺寸图。（略）

附件 3：三种类型的光伏电池（A 单晶硅、B 多晶硅、C 非晶硅薄膜）组件设计参数和市场价格。（略）

附件 4：大同典型气象年气象数据。特别注意：数据库中标注的时间为实际时间减 1 小时，即数据库中的 11:00 即为实际时间的 12:00。（略）

附件 5：逆变器的参数及价格。（略）

附件 6：可参考的相关概念。（略）

附件 7：小屋的建筑要求。（略）

[评阅要点]

本题评阅时请注意：建模的准备工作，数学模型的建立，求解方法及过程、结果的表述、图示、分析，第三问的创新性。

建模的准备工作：这部分是建模及解答的基础（集中或分散描述）。

①倾斜面总辐射强度的计算。这里涉及太阳时、时角、赤纬角、太阳高度角和太阳方位角等概念，还需要了解斜面的阳光直射辐射强度与散射辐射强度的计算。

②附件 4 提供的辐射强度是离散数据，需要将数据连续化，计算光照辐射量。

问题 1：只考虑贴附安装方式建立数学模型及求解。

①建立模型：单目标模型或多目标模型，可考虑发电总量、单位发电费用、经济效益和投资回收年限等。希望学生能够全面地分析问题，建立相应的优化模型。

②模型求解：要求给出求解方法的详细描述。

③结果表述：结果的表述、分析要清楚、明确，要求给出电池铺设图及所配用的逆变器列表。例如，一种可行的铺设方式是：顶面可铺设 40 多块 A3 电池，南面可铺设 30 多块 C10 电池，西面可铺设 C1 和 C10 电池各 10 多块，东、北两面可不铺设（此例只是可行方案之一）。

问题 2：考虑架空安装方式建立数学模型及求解。

① 重点考虑屋顶上架空的光伏电池平面的最佳倾角，约为30度。

② 其他要求同问题1。

问题3：小屋设计。

① 需要计算南墙的最佳朝向，光伏电池板的最优铺设。

② 根据计算结果，设计“最佳”小屋。

③ 本问题重点是考查学生的创造性。

(3)全国赛C题

脑卒中发病环境因素分析及干预

脑卒中(俗称脑中风)是目前威胁人类生命的严重疾病之一，它的发生是一个漫长的过程，一旦得病就很难逆转。这种疾病的诱发已经被证实与环境因素(包括气温和湿度)之间存在密切的关系。对脑卒中的发病环境因素进行分析，其目的是为了进行疾病的风险评估，对脑卒中高危人群能够及时采取干预措施，也让尚未得病的健康人或者亚健康人了解自己得脑卒中风险程度，进行自我保护。同时，通过数据模型的建立，掌握疾病发病率的规律，对于卫生行政部门和医疗机构合理调配医务力量、改善就诊治疗环境、配置床位和医疗药物等都具有实际的指导意义。

数据(见Appendix－C1)来源于中国某城市各家医院2007年1月至2010年12月的脑卒中发病病例信息以及相应期间当地的逐日气象资料(Appendix－C2)。请你们根据题目提供的数据，回答以下问题。

①根据病人基本信息，对发病人群进行统计描述。

②建立数学模型研究脑卒中发病率与气温、气压、相对湿度间的关系。

③查阅和搜集文献中有关脑卒中高危人群的重要特征和关键指标，结合①和②中所得结论，对高危人群提出预警和干预的建议方案。

附件：Appendix-C1、Appendix-C2。(略)

[评阅要点]

本题评阅时请注意：建模的准备工作(包括缺失和误差数据的处理)，模型的表达、求解和分析方法，结果的表述、解释及图示，注重模型的合理性分析及模型的拓广。本题的难点在于是否将脑卒中发病(人数)与环境因素(年龄、气温、气压、相对湿度等)联系起来，建立合适的数学模型，并用于对高危人群进行预警和干预。

问题1：

① 统计出发病人数在病人基本信息(包括性别、年龄段、职业及其他)中的分布规律(如：百分比)，应说明缺失和误差数据的处理。

②分析发病率随年龄的变化规律(如：近似偏正态分布)。

注:除了简单的统计描述或图形外,应有统计规律(可分情况)的提取及其理由的陈述或分析。

问题 2:

① 将统计出的发病人数作为因变量,气温(气温差)、气压、相对湿度作为自变量建立统计回归模型(如:全变量的多元线性回归模型),计算并报告模型中的参数估计、模型拟合误差、预测结果以及显著性变量等。

② 应在 2.1 获得的显著变量基础上,考虑建立单因子统计模型(如:单因子线性、二次回归分析)、单因子方差分析等。

③ 应考虑建立条件统计模型(如:分别对男,女,农民,60≤年龄≤80,发病人数≥60 的情况建立线性回归模型)并进行相应分析讨论。

④ 应考虑异常值识别或剔除,模型的合理性,模型的检验或拓广。

注:①中模型所用样本是按天的,应有模型拟合误差和预测的结果或分析;除了①外,在②③④中应有适当的工作,尤其是②③。

问题 3:

① 查阅文献资料,脑卒中的高危人群重要特征(危险因素)。如:高血压(最危险因素)、心脏病、短暂性脑缺血发作、糖尿病、高血脂、超重与肥胖、吸烟、长期过度饮酒、高盐(偏咸)饮食、缺少运动、性格(争强好胜的 A 型性格)、不可改变因素(性别、年龄、遗传等)以及诱发因素(过度紧张、激动、兴奋、愤怒和疲劳等)。

② 关键指标(定量):查文献给出上述(可测量)重要特征的定量指标(如:血压值、短暂性脑缺血发作次数、血糖值、血脂值、体重指数值、每日吸烟支数、每周饮酒次数、每日平均食盐量、每周锻炼次数、年龄等)。

③ 利用所建立的模型,预先以适当方式具体告诉或提醒具备脑卒中高危人群重要特征和关键指标的当事人在怎样的环境下(如:什么时间、年龄、气压、温度)最容易发病或发病率明显增长,并提出建议方案(如:发病人数与平均相对湿度呈负相关关系;在气压大于一定数值时,发病人数随气压升高而有增大;年龄在 60 到 80 的人群容易发病。对于上述环境因素,高危人群应注意防范)。

注:③的结论应根据问题 1 和 2 所得到的模型的结果来获得,应有量化结果。

(4)全国赛 D 题

机器人避障问题

图 11-3 是一个 800×800 的平面场景图,在原点 $O(0, 0)$ 处有一个机器人,它只能在该平面场景范围内活动。图中有 12 个不同形状的区域是机器人

不能与之发生碰撞的障碍物，障碍物的数学描述如下表：

编号	障碍物名称	左下顶点坐标	其他特性描述
1	正方形	(300, 400)	边长 200
2	圆形		圆心坐标(550, 450)，半径 70
3	平行四边形	(360, 240)	底边长 140，左上顶点坐标(400, 330)
4	三角形	(280, 100)	上顶点坐标(345, 210)，右下顶点坐标(410, 100)
5	正方形	(80, 60)	边长 150
6	三角形	(60, 300)	上顶点坐标(150, 435)，右下顶点坐标(235, 300)
7	长方形	(0, 470)	长 220，宽 60
8	平行四边形	(150, 600)	底边长 90，左上顶点坐标(180, 680)
9	长方形	(370, 680)	长 60，宽 120
10	正方形	(540, 600)	边长 130
11	正方形	(640, 520)	边长 80
12	长方形	(500, 140)	长 300，宽 60

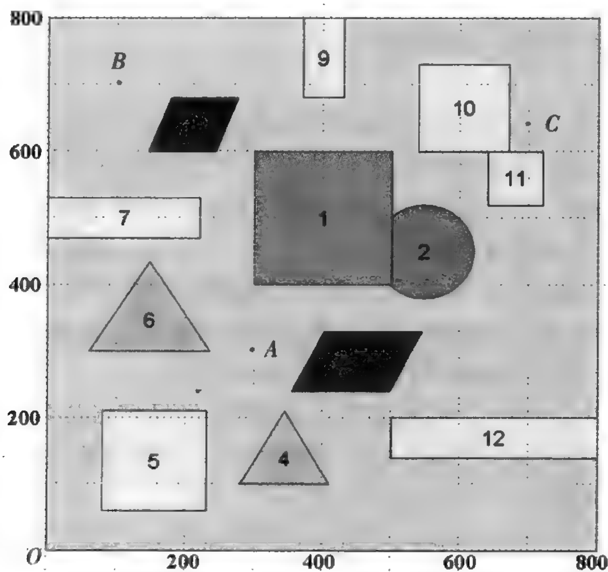


图 11-3 800×800 平面场景图

在图 11-3 的平面场景中，障碍物外指定一点为机器人要到达的目标点（要求目标点与障碍物的距离至少超过 10 个单位）。规定机器人的行走路径由直线段和圆弧组成，其中圆弧是机器人转弯路径。机器人不能折线转弯，转弯路径由与直线路径相切的一段圆弧组成，也可以由两个或多个相切的圆弧路径组成，但每个圆弧的半径最小为 10 个单位。为了不与障碍物发生碰撞，同时要求机器人行走线路与障碍物间的最近距离为 10 个单位，否则将发生碰撞，若碰撞发生，则机器人无法完成行走。

机器人直线行走的最大速度为 $v_0=5$ 个单位/秒。机器人转弯时，最大转弯速度为 $v=v(\rho)=\dfrac{v_0}{1+\mathrm{e}^{10-0.1\rho^2}}$，其中 ρ 是转弯半径。如果超过该速度，机器人将发生侧翻，无法完成行走。

请建立机器人从区域中一点到达另一点的避障最短路径和最短时间路径的数学模型。对场景图中 4 个点 $O(0, 0)$，$A(300, 300)$，$B(100, 700)$，$C(700, 640)$，具体计算：

①机器人从 $O(0, 0)$ 出发，$O\to A$，$O\to B$，$O\to C$ 和 $O\to A\to B\to C\to O$ 的最短路径；

②机器人从 $O(0, 0)$ 出发，到达 A 的最短时间路径。

注：要给出路径中每段直线段或圆弧的起点和终点坐标、圆弧的圆心坐标以及机器人行走的总距离和总时间。

［评阅要点］

问题常见于计算机图形处理和机器人自动行走，其中二维和三维避障问题是热点课题。固定场景中的避障是其中较简单的情形，不涉及障碍物的移动和障碍物的识别。本问题的数学模型和求解算法分量基本相当。

① 预处理：将原始障碍物作适当扩大，确保不发生碰撞。在合理的假设下，要求预处理步骤清晰，说明适当，明确给出最后的处理结果。

② 数学模型：必须明确建立路程最短和时间最少的优化模型，要给出明确的决策变量、目标函数和约束条件，表述准确全面。

③ 求解方法：机器人路径规划求解方法较多，尽量选择合适的求解方法和相应的算法，给出求解方法的思想、求解算法的步骤或流程图，并给出图形的预处理过程和描述。

④ 结果与分析：要有明确的数值结果和路线图，表达简明、清晰。最优路径应由直线段与圆弧光滑连接而成。例如，第一问的圆弧半径为 10，第二问的圆弧半径需通过优化模型计算得到。例如，到 B 点距离最短的最优路径可参见图 11-4，到 A 点时间最短的最优路径可参见图 11-5。

⑤ 深入分析：本问题有较大的发挥空间，参赛者利用参考文献作深入分析时，可考虑适当加分。

讨论不同算法对两种最优结果的影响；

适当考虑直线行走与转弯之间时间延迟的影响；

适当扩充，讨论障碍物变形或不规则障碍物对最优结果的影响等；

适当扩充，定性讨论障碍物的适当移动或目标点的适当移动对结果的影响，或者定性讨论三维空间的情况等。

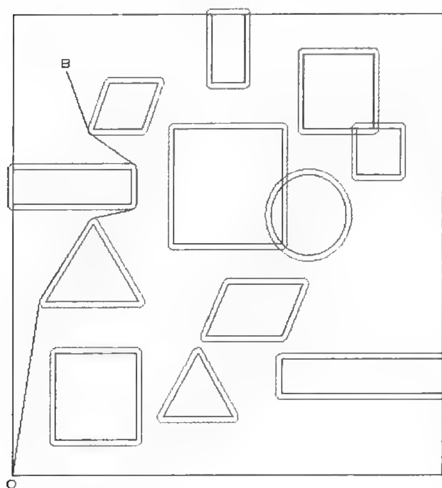


图 11-4

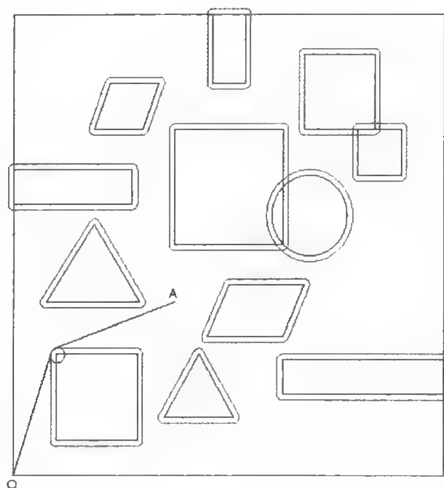


图 11-5

参考文献

[1] 陈晨. 科技手段解决文物鉴定诚信危机[J]. 中国品牌与防伪,2011(7).
[2] 窦志铭. 连锁经营管理理论与实务(第2版)[M]. 北京:中国人民大学出版社,2012.
[3] 方晓,阚燕. 基于情侣博弈的思考[J]. 现代商贸工业,2011(10).
[4] 傅吾豪,著. 袁志强,编. 操盘圣经[M]. 北京:当代中国出版社,2007.
[5] 葛剑雄. 中国人口发展史[M]. 福州:福建人民出版社,1991.
[6] 龚劲涛,吴英. 两种求铅球最佳出手角的方法比较[J]. 物理与工程,2011(3).
[7] 胡良剑,孙晓君. MATLAB数学实验[M]. 北京:高等教育出版社,2006.
[8] 胡向东,魏琴芳. 应用密码学教程[M]. 北京:电子工业出版社,2005.
[9] 胡运权. 运筹学教程[M]. 北京:清华大学出版社,2007.
[10] 黄文璋. 莎士比亚新诗真伪之鉴定[J]. 中国统计,1999(7).
[11] 姜启源,谢金星,叶俊. 数学模型(第4版)[M]. 北京:高等教育出版社,2011.
[12] 蒋新国. 现代药物动力学[M]. 北京:人民卫生出版社,2011.
[13] 金慧娟,吕荷莉,陈志强. 第29届奥运会男子篮球比赛前8名队运动员身体形态特征的比较研究[J]. 中国体育科技,2009(3).
[14] 李凌云. 运动生物力学原理在武术运动中的应用[D]. 济南:山东师范大学,2002.
[15] 李阳. 兰州拉面产业的发展和面临的制约——“马兰”商标异议案引发的思考[J]. 中华商标,2010(2).
[16] 刘占锋. 影响铅球飞行距离的运动生物力学因素分析[J]. 经营管理者,2010(15).
[17] 陆爱云. 运动生物力学[M]. 北京:人民体育出版社,2010.
[18] 马飞. 数学三十六计[M]. 北京:金盾出版社,2008.
[19] 马赞甫,刘妍珺. 基于斗鸡博弈的红绿灯时间设置[J]. 公路交通科技,2011(11).
[20] 齐济. 涨跌易断[M]. 广州:南方日报出版社,2011.

[21] 任伟. 现代密码学[M]. 北京:北京邮电大学出版社,2011.
[22] 阮晓青,周义仓. 数学建模引论[M]. 北京:高等教育出版社,2005.
[23] 孙惠泉. 图论及其应用[M]. 北京:科学教育出版社,2004.
[24] 王青建. 数学开心辞典[M]. 北京:科学出版社,2008.
[25] 王树禾. 数学模型基础[M]. 合肥:中国科学技术大学出版社,1996.
[26] 韦美雁. 赛程安排的数学模型[J]. 宜春学院学报,2007(2).
[27] 熊蓓. 企业竞争的博弈分析[J]. 现代商贸工业,2011(10).
[28] 徐国祥. 统计预测与决策(第 3 版)[M]. 上海:上海财经大学出版社,2011.
[29] 杨攀,杨斌,罗明奎,等. 饮酒驾车的数学模型[J]. 大学数学,2005(6).
[30] 杨启帆,谈之奕,何勇. 数学建模(第 3 版)[M]. 杭州:浙江大学出版社, 2010.
[31] 尹军,李鸿江. 世界优秀男女中长跑运动员身体形态与机能特征的研究[J]. 西安体育学院学报,2004(3).
[32] 张维迎. 博弈论与信息经济学[M]. 上海:上海人民出版社,2004.
[33] 赵静,但琦. 数学建模与数学实验(第 3 版)[M]. 北京:高等教育出版社,2008.
[34] 脂砚斋重评石头记(庚辰本)[M]. 北京:北京图书馆出版社,2004.
[35] Avinash K Dixit,Barry J Nalebuff. 妙趣横生博弈论[M]. 董志强,王尔山,译. 北京:机械工业出版社,2009.
[36] Livi Bacci Massimo. 繁衍:世界人口简史(第 3 版)[M]. 郭峰,庄瑾,译. 北京:北京大学出版社,2005.
[37] Mark M Meerschaert. Mathematical Modeling(第 3 版)[M]. 北京:机械工业出版社, 2009.

[General Information]
书名=生活中的数学模型
作者=管宇主编
页数=223
SS号=13317767
DX号=
出版日期=2013.05
出版社=浙江工商大学出版社

封面
书名
版权
前言
目录
第1章　从山崖测量说起——什么是数学模型
1. 1山崖测量
1. 2数学模型和数学建模
1. 3数学模型分类和数学建模过程
第2章　生活中的数学模型
2. 1桌子一定能放平吗
2. 2雨中行走
2. 3双层玻璃的功效
2. 4连锁经营
第3章　体育中的数学模型
3. 1铅球投掷
3. 2运动生物力学
3. 3体育项目与身体形态特征
3. 4赛程安排
第4章　真伪鉴别
4. 1《红楼梦》前80回和后40回的作者之谜
4. 2统计检验
4. 3莎士比亚新诗真伪之鉴定
4. 4油画造假与鉴定
4. 5文物鉴定的科技方法
4. 6多管齐下才能更可靠
第5章　社会历史演变规律——人口模型
5. 1马尔萨斯人口理论
5. 2逻辑斯蒂克模型
5. 3中国历史人口演变
5. 4世界历史人口演变
5. 5更精细的人口模型
第6章　饮酒、吃药和打针——动力学模型
6. 1饮酒驾车
6. 2房室模型
6. 3饮酒驾车的数学模型
6. 4传染病问题

第7章　战争与和平——博弈理论
7.1什么是博弈
7.2现代博弈理论
7.3博弈论的基本概念
7.4博弈案例
第8章　图论与网络分析
8.1基础知识简介
8.2哥尼斯堡七桥问题与欧拉图
8.3匹配问题
8.4中国邮递员问题
8.5游山玩水问题
第9章　密码学及其应用
9.1密码故事
9.2密码学发展简史
9.3密码学基本知识
9.4代替密码
9.5　RSA密码体制
9.6密码学其他案例
第10章　股市实战——预测与决策
10.1股票和股票交易
10.2波段操作
10.3技术指标分析
10.4旁门左道
10.5预测与决策
第11章　数学建模竞赛
11.1数学建模竞赛
11.2数学建模竞赛题目
参考文献